**Jarno Hänninen**

# The effect of selective attention on auditory frequency selectivity in humans

**School of Electrical Engineering**

Thesis submitted for examination for the degree of Master of Science in Technology.

Espoo October 10, 2016

**Thesis supervisor:**

Prof. Mikko Sams

**Thesis instructors:**

D.Sc. (Tech.) Jaakko Kauramäki

Ph.D. Iiro Jääskeläinen

Aalto University
School of Electrical
Engineering

AALTO UNIVERSITY
SCHOOL OF ELECTRICAL ENGINEERING

ABSTRACT OF THE
MASTER'S THESIS

Author: Jarno Hänninen

Title: The effect of selective attention on auditory frequency selectivity in humans

Date: October 10, 2016 Language: English Number of pages:8+83

Department of Signal Processing and Acoustics

Professorship: Acoustics and Audio Signal Processing Code: S3004

Supervisor: Prof. Mikko Sams

Instructors: D.Sc. (Tech.) Jaakko Kauramäki, Ph.D. Iiro Jääskeläinen

When we listen to a friend in a noisy café, the segregation between information and noise depends largely on the differences in neural level representations of the auditory inputs. The auditory cortex (AC) is tonotopically organized, which means that the frequency information is mapped to the cortical topography. Thus, similar frequencies are processed in close vicinity at AC. The tonotopy set limits to the frequency resolution, but under selective attention the frequency-based segregation succeeds even in conditions where the signal and noise are overlapping in frequency. The aim of the study was to explore these underlying short-term neural mechanisms that can sharpen the frequency selectivity at AC under selective attention.

The topic was investigated by a psychophysical experiment where parametric changes in attention and background noise were used to bias the cortical responses. In the experiment auditory evoked magnetic field components N100m (100–200 ms post-onset) and SF (200–600 ms post-onset) were measured from 14 subjects with MEG during a behavioral task. In the behavioral task subjects identified 1020-Hz target tones (P=0.1) from more frequently occurring 1000-Hz standard tones (P=0.9), or respectively focused their attention on a visual control task where identical auditory stimuli were playing. The magnetic fields were evoked by a sequence of 300-ms pure tones that were suppressed with continuous-time notched-noise maskers. The difficulty level of the auditory task (7 levels) was varied by changing the band-stop filtered spectral "notch" around the target stimuli between $\pm$500 Hz and 0 Hz.

The current data showed that the N100m and SF components were stronger under selective auditory attention. The impact was most robust at 200–500 ms post-onset when the notches were within the critical band ($\leq \pm$160 Hz). Moreover, the effect was more pronounced in the left auditory cortex, especially with the thinnest notches. The results indicate that the neurophysiological mechanisms of selective auditory attention are based on neural gain when the notches are clearly wider than the critical band and on neural tuning (gain + selectivity increase) when the notches are within the critical band. Moreover, it seems that the left auditory cortex has a more active role in conditions where the segregation of relevant sounds from noise requires very sharp filtration.

Keywords: auditory cortex, frequency selectivity, selective attention, magnetoencephalography, auditory-evoked fields, neural tuning

AALTO-YLIOPISTO
SÄHKÖTEKNIIKAN KORKEAKOULU

DIPLOMITYÖN
TIIVISTELMÄ

Tekijä: Jarno Hänninen

Työn nimi: Valikoivan tarkkaavaisuuden vaikutus hermosolujen auditoriseen taajuusselektiivisyyteen ihmisillä

Päivämäärä: October 10, 2016 Kieli: Englanti Sivumäärä:8+83

Signaalinkäsittelyn ja akustiikan laitos

Professuuri: Akustiikka ja äänenkäsittely Koodi: S3004

Valvoja: Prof. Mikko Sams

Ohjaajat: TkT Jaakko Kauramäki, PsT Iiro Jääskeläinen

Kun auditorinen tarkkaavaisuus kohdistetaan tiettyyn äänen ominaisuuteen (esim. sävelkorkeus) taustamelun aikana, ihmisen kuuloaivokuorella tapahtuu lyhytaikaisia neuraalisia muutoksia, jotka tehostavat hyödyllisen informaation erottelua kohinasta. Tutkimuksen tarkoituksena oli kartoittaa kuuloaivokuoren neuraalisia mekanismeja, jotka parantavat kuulon taajuus-selektiivisyyttä erityisesti vaativissa kohinaolosuhteissa.

Tutkimuksessa selvitettiin kuuloaivokuoren neuraalista toimintaa psykofyysisellä kokeella, jossa mitattiin magneettisia N100m (latenssi 100–200 ms) ja SF (latenssi 200–600 ms) herätevasteita MEG-laitteella 14 koehenkilöltä behavioraalisen tehtävän aikana. Behavioraalisessa tehtävässä koehenkilöt erottelivat 1020-Hz merkkiääniä (P=0.1) yleisemmistä 1000-Hz standardiäänistä (P=0.9) tai tekivät visuaalista kontrollitehtävää ääniärsykkeiden soidessa taustalla. Herätteinä käytimme 300 ms siniääniä, joiden havaitsemista vaikeutettiin jatkuva-aikaisilla kohinamaskereilla. Kohinamaskerit (7 eri vaikeustasoa) luotiin suodattamalla valkoiseen kohinaan vaihtelevia päästökaistoja 1 kHz ympärille $\pm$500 Hz ja 0 Hz väliltä. Kokeellinen paradigma mahdollisti auditorisen tarkkaavaisuuden ja taustakohinan neuraalisten vaikutusten analysoimisen vertailemalla neuraalista ja behavioraalisia vasteita keskenään.

Mittaukset osoittivat, että N100m ja SF herätevasteet olivat voimakkaampia selektiivisen auditorisen tarkkaavaisuuden aikana kuin visuaalisessa kontrollitilanteessa. Vahvistuminen oli voimakkainta vaativissa kohinaolosuhteissa – kun päästökaista oli kriittistä kaistaa ($\leq \pm$160 Hz) kapeampi latenssien ollessa 200–500 ms. Yllättäen tämä neuraalisen aktiviteetin kasvu oli voimakkaampaa vasemmalla aivopuoliskolla. Tulokset osoittavat, että neurofysiologinen tarkkaavaisuusmekanismi on gain-tyyppistä, kun kohinan päästökaista on reilusti kriittistä kaistaa leveämpi. Kohinan päästökaistan ollessa kriittistä kaistaa kapeampi tulokset tukevat tuning-tyyppistä mekanismia (gain + selektiivisyyden kasvu). Vaativimmissa kohinaolosuhteissa selektiivisyyden kasvulla näyttää olevan suurempi merkitys, joka korostuu erityisesti vasemman aivopuoliskon vasteissa.

Avainsanat: kuuloaivokuori, taajuusselektiivisyys, selektiivinen tarkkaavaisuus, magnetoenkefalografia, auditorinen herätevaste, hermosolujen ärsykepiirrevirittys

# Preface

This master's thesis was done in the Department of Biomedical Engineering and Computational Science (BECS) of the Aalto University School of Science and Technology. The supervisor of this work was Prof. Mikko Sams and the instructors D.Sc. (Tech.) Jaakko Kauramäki and Ph.D. Iiro Jääskeläinen.

I would like to thank my both instructors Jaakko and Iiro, as well as my supervisor Mikko for the opportunity to work in the BECS and the encouragement you have given given me throughout this period. I would also like to thank my co-workers and friends who took part in the psychophysical experiment.

Most of all, I would like to thank my family members and my parents for the support they have given me throughout my studies. I would like to thank my wife Sini for her love and patience. Special thanks to my three-year-old boy who always asks unexpected questions. Sini and Akseli, I love you so much.

Espoo, October 10, 2016

Jarno L. J. Hänninen

# Contents

# Symbols and abbreviations

| | |
|---|---|
| $\mathbf{B}$ | magnetic flux density |
| $\mathbf{E}$ | electric field density |
| $f$ | frequency |
| $I$ | electric current |
| $\mathbf{J}$ | current density |
| $k$ | Weber's constant |
| $p$ | pressure |
| $P$ | probability |
| $q$ | electric charge |
| $\mathbf{Q}$ | electric dipole moment |
| $t$ | time |
| $V$ | electric potential |
| $\overline{x}$ | arithmetic mean |

| | |
|---|---|
| $\epsilon_0$ | electric permittivity of free space |
| $\lambda$ | length constant of the membrane |
| $\mu_0$ | magnetic permeability |
| $\rho$ | free electric charge density |
| $\sigma$ | conductivity |
| $\sigma$ | standard deviation |

| | |
|---|---|
| ABR | auditory brainstem response, I–VI |
| AC | auditory cortex |
| AEF/AEP/AER | auditory evoked field/potential/response |
| ANOVA | analysis of variance |
| AttAud | auditory attentive condition |
| AttVis | visual control condition |
| BF | best frequency (a.k.a. CF) |
| BM | basilar membrane |
| CANS | central auditory nervous system |
| CB | critical band |
| CF | characteristic frequency |
| CNS | central nervous system |
| dB | decibel |
| ECD | equivalent current dipole |
| EEG | electroencephalography |
| EOG | electro-oculography |
| EPSP | excitatory postsynaptic potential |
| ERF/ERP/ERR | event-related field/potential/response |

| | |
|---|---|
| fMRI | functional magnetic resonance imaging |
| HPI | head position indicator coils |
| Hz | Hertz, 1/s |
| IC | inferior colliculus |
| IHC | inner hair cell |
| ISI | interstimuli interval |
| IPSP | inhibitory postsynaptic potential |
| JND | just-noticeable difference |
| LLR | long-latency response, P50(m), N100(m) and P200(m) |
| MEG | magnetoencephalography |
| MLR | middle-latency response, P0, Na, Pa and Nb |
| MRI | magnetic resonance imaging |
| MSR | magnetically shielded room |
| N100/N100m | negative peak in the AEP/AEF with latency of around 100 ms |
| OHC | outer hair cell |
| P50/P50m | positive peak in the AEP/AEF with latency of around 50 ms |
| P200/P200m | positive peak in the AEP/AEF with latency of around 200 ms |
| PAC | primary auditory cortex |
| PET | positron emission tomography |
| PFC | prefrontal cortex |
| PPC | posterior parietal cortex |
| PSD | power spectral density |
| SEM | standard error of the mean |
| SF/SP/SR | sustained field/potential/response |
| SNR | signal-to-noise ratio |
| SPL | sound pressure level |
| SQUID | superconducting quantum interference device |
| STG | superior temporal gyrus |
| T | tesla, 10 000 G (gauss) |

# Chapter 1

# Introduction

One side effect, which relates to the easiness to produce and distribute music with digital techniques is the oversupply of music. In public places we are forced to hear auditory filth that is composed by non-professionals – typically an adman or economist, and performed by persons with lack of self-criticism[1]. On top of that, the auditory stream (filth) that we receive continuously from the PA speakers and portable gadgets, is masked by urban traffic noises. Although we suffer from this cacophony, our auditory system can segregate the information-bearing patterns (signal) from random patterns (noise). Nonetheless, evolution has provided us with powerful neural mechanisms with which we can attend and ignore sensory events and interact efficiently in the noisy world around us. By means of attention, we can navigate coherently in urban traffic, chat with friends in a crowded café, and enjoy the delicate melodic structure of a polyphonic music piece.

At the time of epistemic 'Copernican Revolution' Immanuel Kant claimed in *The Critique of Pure Reason*[2] that perception is not a one-way process where information flows from the object to the deterctor, but rather an interactive process where the detector adapts to the object (Kant and Kehrbach, 1910). Another revolutionary concept that he manifested was a theory of *schema*. Schema is a mental representation of a specific physical event, which is constantly compared to the information that is received from the environment. When a physical event matches to the corresponding schema, also the semantically relevant mental systems activate and cause psychological reactions.

The Kantian concepts from the 18th century Enlightenment are still useful in the modern neuroscience, especially in the research of auditory attention. When we perceive a sound, the pressure oscillations of the air set the inner structures of the ear to vibrate, then this unique vibration pattern that depends on the physical properties of sound is transduced to the neural responses that are transmitted to the auditory cortex. The auditory cortex represents the physical properties of a sound – intensity, frequency and duration by activation patterns of neurons. These stimulus-related activation patterns can be then

[1] The music scene was of course not optimal in the old times. At that time the beautiful harmonies and polyfonies that were composed by true craftsmen, were exploited to nationalist-religious purposes.

[2] Kritik der reinen Vernunft

studied non-invasively with magnetoencephalography (MEG) or electroencephalography (EEG), which gives a possibility to study neural functions indirectly (stimulus-responce approach).

The *cocktail party example* (Cherry, 1953) demonstrates the function of selective auditory attention with dynamic filters that pass the relevant auditory stream and attenuate the non-relevant streams. The filtration is based largely on frequency-based selection, so that only the thin frequency band around the target signal is passed and a vast majority of the background noise is cut-off. The resolution of frequency-based selectivity depends on the physiology of the auditory system and especially on how it processes the spectral information. Complex sounds are decomposed to sinusoidal components in the inner ear's cochlea, and the component frequency is then encoded to the cortical topography. This tonotopy means that similar frequencies are processed in close vicinity at auditory cortex. Thus, the discrimination is easy when the signal and noise differ from the frequency, but when the frequencies are close – within the critical band – auditory inputs tend to stimulate identical receptive fields at auditory cortex and cause perceptual interference called auditory masking.

The tonotopical nature of the auditory system set limits to the frequency selectivity, but selective attention can modulate the processing of neural ensembles at auditory cortex, which increases the selectivity of attended frequency. In the current study we tried to investigate the underlying neurophysiological mechanisms that are crucial when frequency based information is segregated under difficult noise conditions.

# Chapter 2

# Background

## 2.1 Human auditory system

The *human auditory system* detects auditory events by capturing the pressure oscillations of the air with mechanical lever system, and by tranducing this mechanical energy into electrochemical responses of neurons. The physical attributes of a sound event – frequency, intensity and duration are encoded to the activation patterns of neural ensembles at cerebral auditory cortex. These activation patterns then reflect the perceived auditory events, which gives an unique identity to the auditory object, such as recognizable timbre of a musical instrument or a familiar voice of a friend. The capability of the human auditory system to identify auditory objects according to pitch, timbre and loudness[1] and localize them in space is remarkable, and is based widely on, how the auditory inputs are processed in the central nervous system. The sophisticated processing capabilities ensure that the human auditory system can generate an accurate neural representation of the acoustical landscape, which is also important in the viewpoint of survival (Ashmore, 2008; Karjalainen, 2009; Rossing et al., 2002).

### 2.1.1 The sensitivity of hearing

The auditory organs can detect sounds that are barely greater than the natural molecular movement of air ($\sim 10^{-12}\ W/m^2$) but can also handle high intensities ($\sim 10\ W/m^2$) without instant damage. The intensity ratio between the two extremities can be even $10^{13}$ and thus, it is practical to describe the extensive dynamic range with logarithmic scale. The sound pressure level (SPL) contrast the intensities to the fixed reference level $p_0$ (20 mPa), which is the intensity at the *threshold of hearing* (1 kHz). The sound pressure levels are typically expressed in decibel scale (dB):

[1] Pitch, timbre and loudness are subjective qualities, whose physical correlates are frequency, harmonic compound and intensity of a sound event.

$$20 \log \frac{p}{p_0} \tag{2.1}$$

, which is practical because of the extreme divergences in the intensity ratios (Rossing et al., 2002).

The frequency range of hearing is roughly 10 octaves[2] and falls between 20 and 20 000 Hz. The absolute threshold curve[3] describes the frequency dependent detection threshold of the average human ear. In the figure (Figure 2.1, p. 4) we can see that the sensitivity of the human auditory system varies according to sound frequency. The human auditory system is most sensitive at frequency range where the majority of speech information occurs ($\sim$ 1–5 kHz), and being less sensitive to low and high frequencies (Rossing et al., 2002; Karjalainen, 2009).

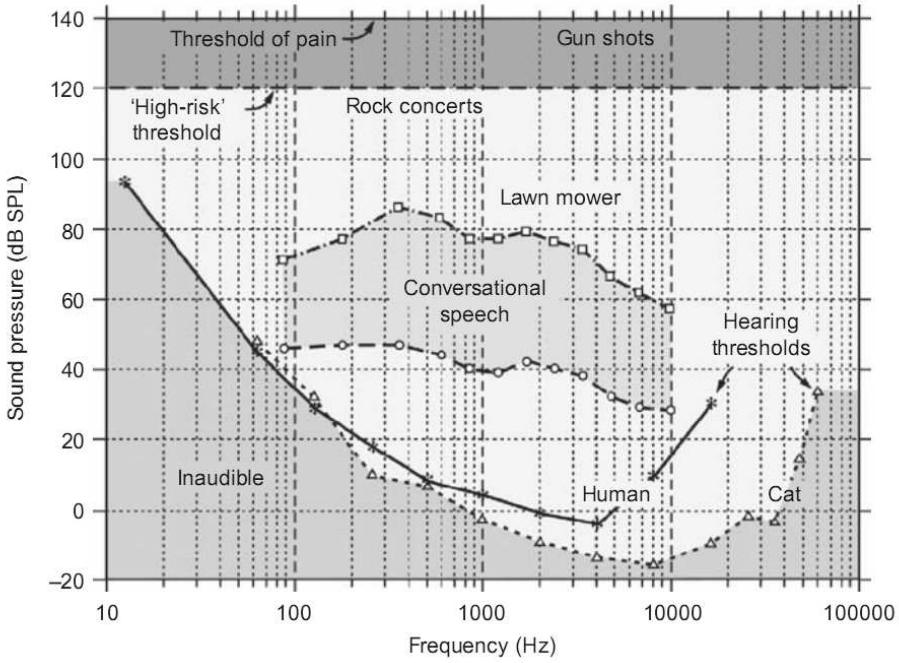



Figure 2.1: Threshold of hearing and dynamic range (adapted from Baars and Gage, 2010).

### 2.1.2 Ear anatomy and physiology

The peripheral auditory system is the initial processing stage of the auditory inputs and has three layers: *outer ear*, *middle ear* and *inner ear* (Figure 2.2, p. 5). The pressure oscillations of the air reach first the *pinna* (a.k.a auricle) that collects, pre-processes and channels audio signal into the *ear canal* (a.k.a auditory canal). The pre-processing of the

[2]The distance between the frequencies 20 Hz and 20480 Hz is 10 octaves ($\log_2 \frac{20480Hz}{20Hz} = 10$)

[3]Also known as the *minimum audible field* (MAF).

pinna helps in sound source localization even though the majority of cues come from the interaural time difference (ITD) and the interaural level difference (ILD)[4] (Karjalainen, 2009).

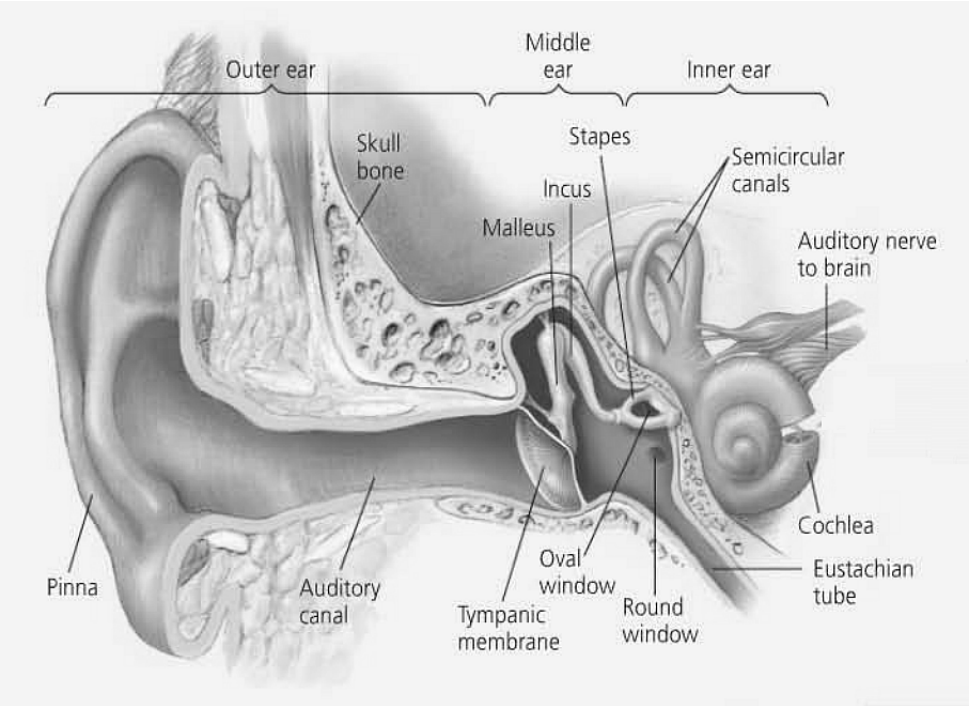



Figure 2.2: The anatomy of ear (adapted from Campbell and Reece, 2009).

The ear canal is a 25 mm long and 7.5 mm wide tube, which is closed at one end. In acoustical sense the ear canal is a closed-end air column. When the pressure wave reaches the bottom of the tube, it sets the elastic *tympanic membrane* (a.k.a *eardrum*) in vibration and reflects back. The ear canal amplifies the vibration of the eardrum about 6 dB (SPL) at its natural frequency ($\sim$ 3–4 kHz), which favours the signals at speech range (Karjalainen, 2009).

The middle ear is an air-filled cavity between outer ear and inner ear containing three tiny bones (Figure 2.2, p. 5) the *malleus*, *incus* and *stapes* (Latin for hammer, anvil and stirrup). This *ossicular chain* (a.k.a *ossicles*) is an ideal mechanism for transmitting oscillation between gases and solids[5], because of the lever factor and it's ability to match the acoustic impedance between mediums[6]. The lever system can amplify the dynamics of the ear (impulses) up to 30 dB (SPL) (Ashmore, 2008).

When the tympanic membrane vibrates the ossicular chain transmits the mechanical energy to the inner ear's fluid-filled *cochlea* through to the membrane in the oval window, so

[4]Interaural level difference is caused by the acoustic shielding of the head.

[5]The ossicular chain converts low pressure and high particle velocity of air to high pressure and low particle velocity of fluid.

[6] The middle ear's ability to match the acoustic impedance between mediums smoothly is based on the lever factor of the ossicular chain and the area ratio between the tympanic membrane and the stapes footplate.

that the impedance is matched optimally at the interface. The oscillation of the cochlear fluid sets the elastic structures of the inner ear in motion stimulating *hair cells* that convert the mechanical vibration pattern into electrical signals. These neural signals are then transmitted and processed in the *central auditory nervous system* (Ashmore, 2008; Karjalainen, 2009).

The inner ear encloses the cochlea (Latin for snail), which is 34 mm long and 2 mm wide spiral-formed cavity containing the auditory receptors, and semicircular canals that compose the organ of balance (Figure 2.2, p. 5). The spiral of cochlea has two and a half turns and it is surrounded by rigid bony walls (Figure 2.3, p. 6). The cochlea has three fluid-filled elongated sections, which run along its length from base to apex: the *vestibular canal* on top, the *cochlear duct* (an elastic structure) in the middle and the *tympanic canal* on the bottom. The pressure waves enter the fluid-filled cochlea through the oval window sets the eleastic cochlear duct with interlinked auditory receptor structures in motion (Figure 2.4, p. 7). A counter opening to the oval window – the round window – allows the incompressible fluid space to expand during the motion. (Alberti, 2001; Brownell and Oghalai, 2009).

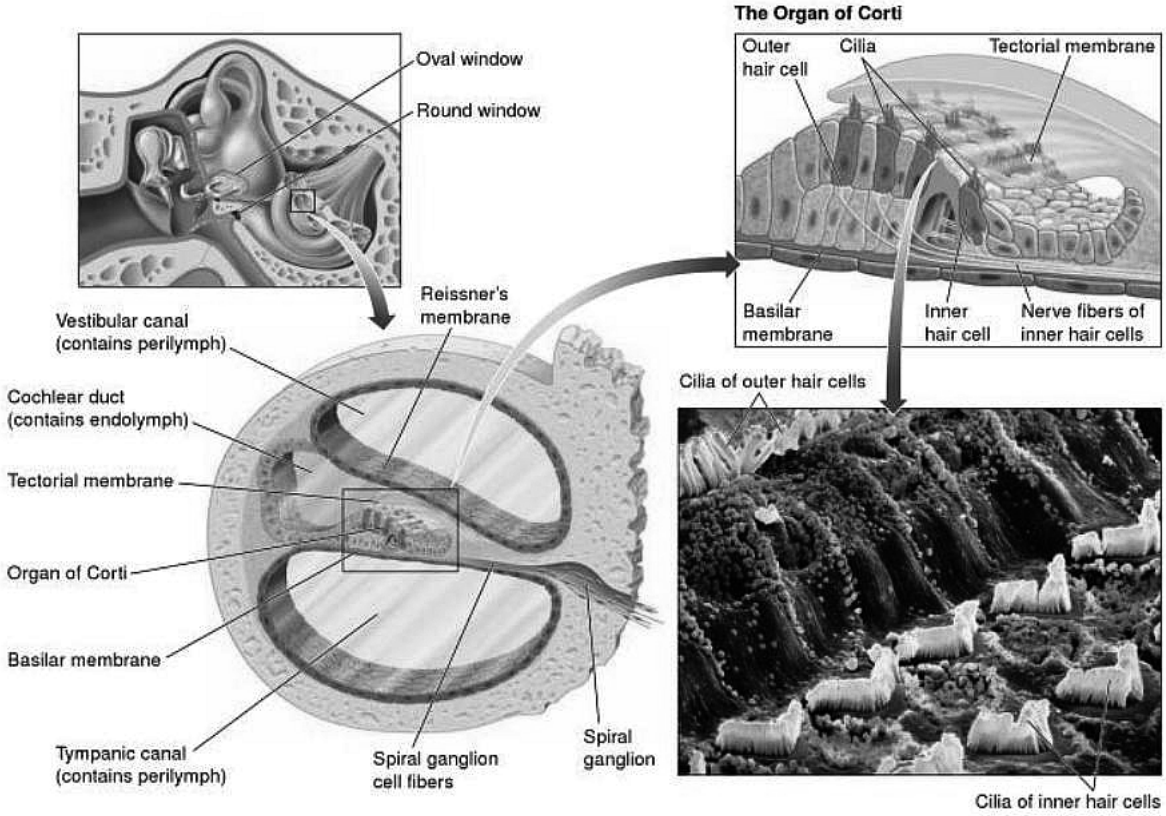



Figure 2.3: Cross-section of the cochlea (adapted from Freberg, 2009).

The cochlear duct is filled with intracellular fluid called *endolymph*, while the other two sections are filled with extracellular fluid called *perilymph*. Perilymph and endolymph have different ionic composition[7], which leads to electric potential that powers the func-

[7]The electric potential of endolymph is $\sim$ 80–90 mV more positive than perilymph due to a higher

tion of receptor cells (Konishi et al., 1978; Purves et al., 2004).

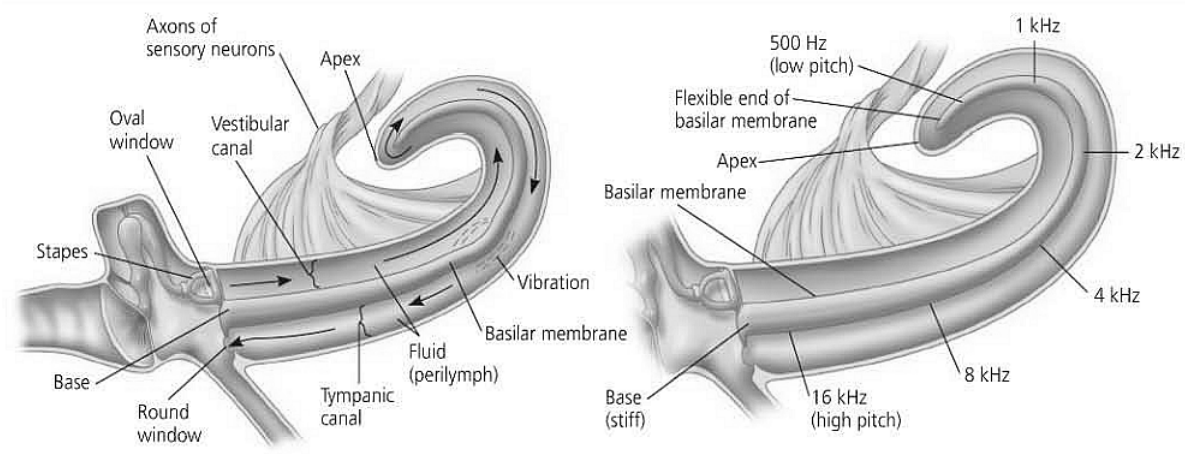



Figure 2.4: The traveling sound wave in the uncoiled cochlea (adapted from Campbell and Reece, 2009).

The *basilar membrane* is an elastic diaphragm within the cochlea that the pressure waves set in motion. The *traveling wave* starts from the base and propagates through the membrane until reaches its maximum displacement and rapidly decays. The shape and stiffness of the membrane varies gradually along its length, so that it is narrow and stiff at the basal end and wide and flexible at the apical end (figure 2.5, p. 8). In the 1930s Hungarian biophysicist Georg von Békésy noticed that the spectral information of the input sound is mapped to the oscillation pattern of the basilar membrane. He found out out that the displacement location on the basilar membrane depends on the frequency (and the intensity) of the incoming sound. High frequencies reach their maximum displacement at the basal end and low frequencies at the apical end of the basilar membrane (Figure 2.4, p. 7). This findind led further to the *place theory of hearing*; the complex wideband sounds are decomposed to the sinusoidal frequency components in inner ears cochlea, and these components are then processed separately at the higher stages of the auditory system (Ashmore, 2008; Karjalainen, 2009).

The mechanical energy is converted to the electrical energy by the *organ of Corti*, which is a cellular layer on the top of basilar membrane (Figure 2.3, p. 6). The organ of Corti has polarized epithelial cells (*hair cells*) on its surface that convert mechanical movements into the elctrical impulses. The hair cells are named according to the hair-like structures *stereocilia* that grow up on the apical surface of the cell (Brownell and Oghalai, 2009). The displacement of the basilar membrane causes shearing force between the hair cell and the overlying *tectorial membrane*, which deflects the stereociliary bundles. The deflection of the stereociliary bundles activate mechanosensitive ion channels that can trigger an electrical impulse (*action potential*; see Section 2.3.1, p. 16) to the *auditory nerve*. The hair cells are organized regularly on the entire length of the basilar membrane. Thus, the frequency information of a sound wave is encoded to the location of the excitated hair

concentration of potassium (K) compared to sodium (Na).

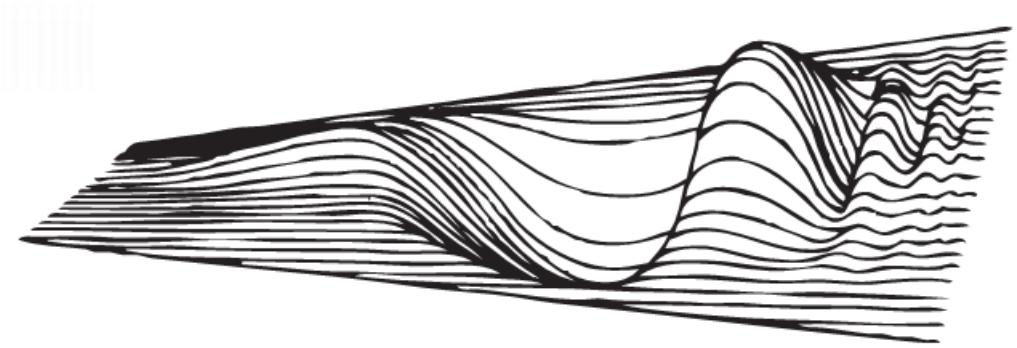


Figure 2.5: The traveling wave on the basilar membrane (adapted from Seikel et al., 2009).

cells. This *tonotopy*[8] is retained also in the higher stages of the human auditory system where the frequencies are mapped to the topography of the neural structures (Ashmore and Gale, 2000; Purves et al., 2004).

The hair cells are not connected straight to the auditory nerve fibers, but they modulate the firing rate of the nerve fibres. There is always a spontaneous discharge rate of action potentials in the auditory nerve fibers (figure 2.6, p. 8). The receptor potential of the hair cell can increase (*excitation*) or decrease (*inhibition*) the discharge rate of action potentials. The excitatory or inhibiry activity depends on the direction in which the stereociliary bundles are bent[9] (Ashmore, 2008; Purves et al., 2004).

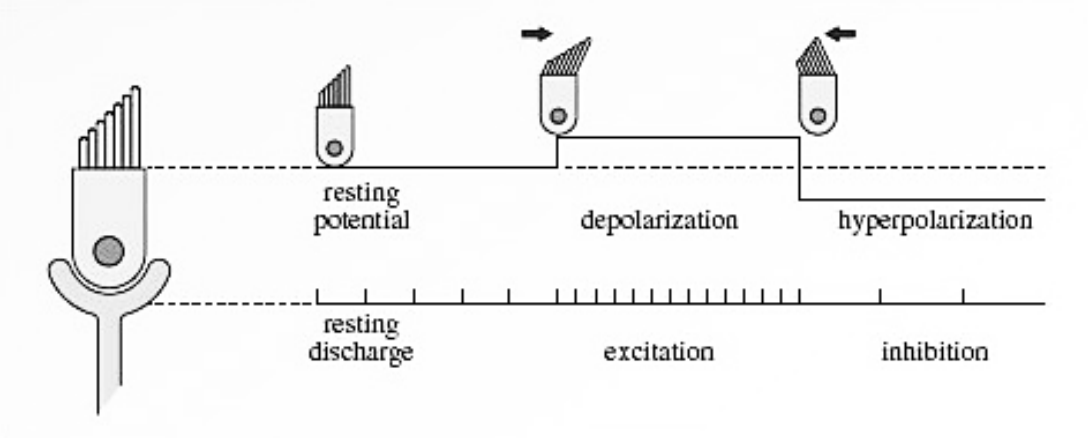



Figure 2.6: The depolarization and hyperpolarization of the hair cell (adapted from Ashmore, 2008).

The frequency, intensity and duration of the auditory stimuli are encoded into a sequence of action potentials, which are then transferred to the auditory nerve. The auditory nerve is a composition of nerve fibers that connect the inner ear to the *central auditory nervous*

[8]The word tonotopy comes from Greek; tono means tone and topos place.

[9]Lateral displacement of cilia towards its tallest edge depolarizes the hair cell and the deflection to the opposite direction closes the channels and the hair cell hyperpolarizes (Ashmore, 2008).

*system* (CANS). The frequency information is place coded and distributed in multiple channels, so that the frequencies close to each other are channeled into same nerve fibers. Each nerve fiber has a specific *characteristic frequency* (CF) – a frequency at which the threshold is lowest and to which it is therefore most sensitive. The neural tuning curves (Figure 2.7, p. 9) represent the minimum sound level in frequency domain that is required to increase a fiber's firing rate above its spontaneous firing rate (Purves et al., 2004).

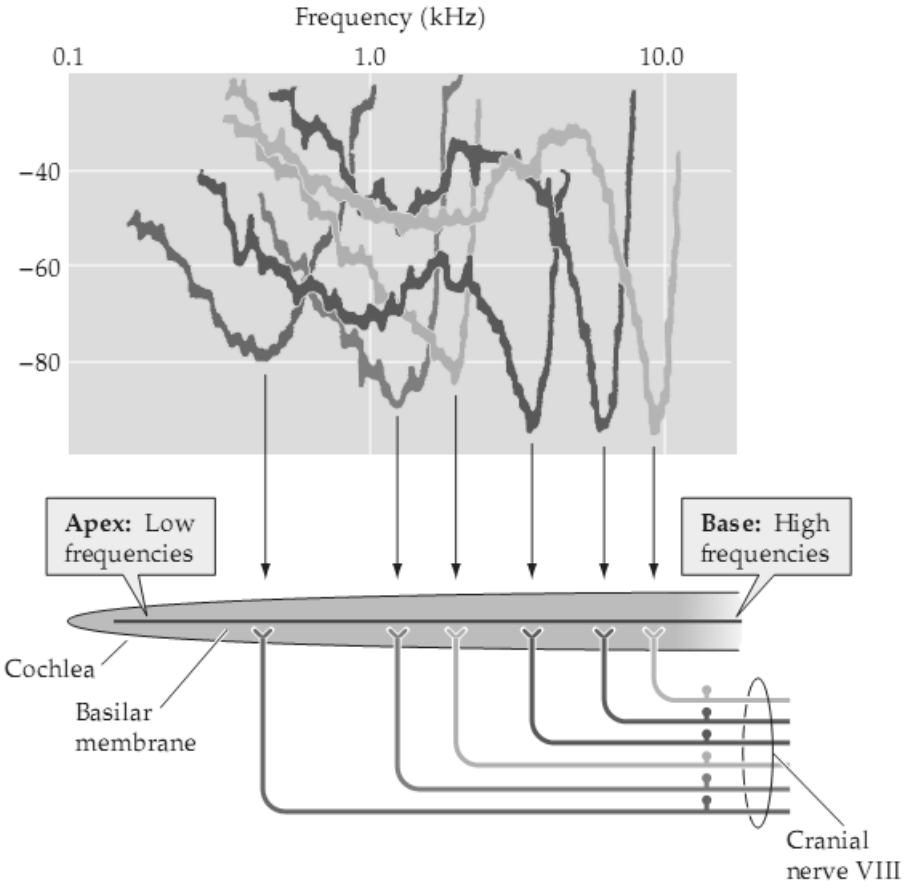



Figure 2.7: The frequency tuning curves of auditory nerve fibers (adapted from Purves et al., 2004).

The place coding is not the only way how the human auditory system encodes frequency information. The firing rate in the auditory nerve fibers carry also information of the frequency, which is referred as *frequency coding* or *phase locking*. When the nerve fibre fires consistently at the same phase of each cycle of a sound wave, the frequency of the sound can be determined from the neuron's firing rate. The significance of the frequency coding is most urgent at low frequencies below 50 Hz. At high frequencies the recovery time between discharges (*refractory period*) is typically 1 ms limits the frequency coding to 1 kHz. The intensity information is encoded to firing rates in the auditory nerve fibers, but also to the amount of excitated fibers. The deflection in the basilar membrane rises the firing rates, but higher intencities cause also larger ares to deflect, which increases the number of nerve fibers to respond the stimulation (Ashmore, 2008).

There are two types of hair cells depending on their function and innervation; about 3500 primary sensory cells, *inner hair cells* (IHCs) and about 9000 to 12,000 secondary sensory cells, *outer hair cells* (OHCs) (Figure 2.3, p. 6). Both hair cell types convert mechanical energy into electrical energy, but only the inner hair cells are connected to the *afferent* nerve fibres that carry nerve impulses from receptors toward the central nervous system. Outer hair cells are connected to the *efferent* nerve fibers that correspondingly bring nerve impulses back from the central nervous system that controls the function of OHCs. Although outer hair cells do not forward auditory information to the higher auditory structures, they have an important function to improve the dynamics and frequency resolution of hearing (Brownell and Oghalai, 2009; Ashmore, 2008).

The electromotile properties of the outer hair cells gives explanation of the wide dynamic range of the human auditory system and is known as *cochlear amplification*. OHCs have a motor protein called prestin on the lateral wall that responds to electrical stimulation by changing the length of the cell. Thus OHCs can convert electrical energy received from cilia excitation back to mechanical energy and enhance the vibration of the organ of the Corti. The cochlear amplification of vibration counteracts viscous damping and thus improve the sensitivity of the human auditory system (Brownell and Oghalai, 2009). The efferent connections that originate from higher auditory layers are suggested to be an *inhibitory feedback system* to control the cochlear amplification. The aforementioned amplification and control system can explain the sensitivity of hearing that can detect mechanical movements at the eardrum that are less than the size of hydrogen atom (Brownell and Oghalai, 2009).

## 2.2 Auditory nervous system

### 2.2.1 Central nervous system

In the previous sections we reviewed how the pressure oscillations of the air are converted into the electrochemical (neural) impulses and how the frequency and intensity information is encoded to these neural impulses. From auditory periphery the neural impulses are transferred to the upper levels of the auditory system, which is part of the *central nervous system* (CNS). The central nervous system is a colossal network of $10^{10}$ *nerve cells* (a.k.a. neurons) with $10^{15}$ *synaptic connections*, and $10^{11}$ *glia cells*. Glia cells provide supporting structure to the neural network by maintaining chemical homeostasis and forming *myelin*, which is the dielectric material on the neuron surface. The information transfer and processing in CNS is based on the dynamics of electrically active neurons that can intermediate electrical signals over long distances with action potentials (more in Section 2.3.1, p. 16) (Martin, 2003; Purves et al., 2004).

Neurons are organized into functional entities (*systems*) that process sensations, perceptions and behaviors. *Sensory systems* acquire and process information from the environment and *motor systems* respond to sensory information by generating movements and behavior. *Associational systems* combine information from sensory system and motor sys-

tems, which is needed in more complex cognitive tasks. Sensory system is divided into sub-sections according to the modality: visual system, *auditory system*, somatosensory system (touch), gustatory system (taste), olfactory system (smell) and vestibular system (balance and movement) (Martin, 2003; Purves et al., 2004).

The central nervous system of humans and other vertebrates comprises the *brain* (*encephalon*) and the *spinal cord* (*medulla spinalis*). The brain floats in the cerebrospinal fluid, which functions as a chemical buffer to protect its fragile structures from impacts. The brain is divided to left and right *cerebral hemispheres* with the sagittal fissure (a.k.a. longitudinal cerebral fissure). The hemispheres are connected together with corpus callosum that consists of 200–250 million contralateral axonal projections (commissures[10]) to ensure fast interhemispheric communication. Other anatomical division of the brain are (Figure 2.8, p. 11): diencephalon, cerebellum, and brainstem. Furthermore, the brainstem has three subdivisions: midbrain, pons and medulla (Purves et al., 2004; Martin, 2003).

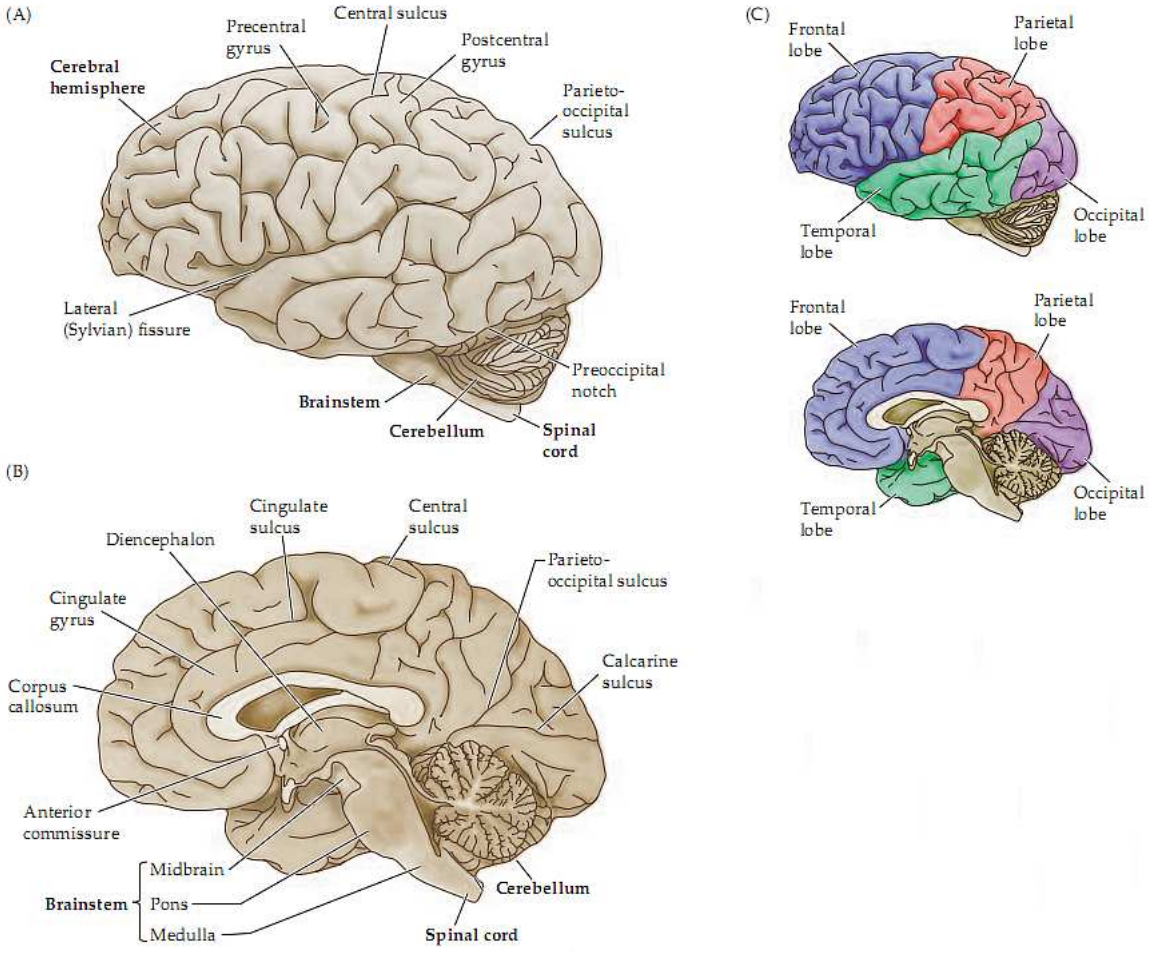



Figure 2.8: The anatomy of the human brain (adapted from Purves et al., 2004).

The majority of neural computing occurs in the outer layer of the cerebral hemispheres in

[10]The tracts that cross the midline of the brain are called as commissures.

the *cortex*. The cortex has a lot of cell bodies and unmyelinated fibers, which gives the gray color to the *gray matter*. The deeper structures of the cortex, the *white matter* consist of myelinated *axons* giving the visual contrast to the gray matter. Axons are assembled into compositions called *tracts* similarly than the axons are composed to nerves in the auditory periphery (Purves et al., 2004; Martin, 2003). The surface area of the cortex is folded, which increases the surface area of the cortex without a notable increase in volume. The cortical folding has enabled new functional areas to evolve for primates and especially for humans to support higher cognitive skills, such as linguistic functions (Barton and Harvey, 2000).

The cortex is typically divided anatomically in four lobes (Figure 2.8, p. 11): frontal, parietal, occipital, and temporal lobes. The lobes are named according to the cranial bones that overlie them (Martin, 2003). The sensory functions are distributed anatomically in such a way that the *primary auditory area* is in *temporal lobe*, *primary visual area* in occipital lobe and *primary somatosensory area* in the parietal lobe (Figure 2.9, p. 12). The primary sensory areas are surrounded by *secondary sensory areas* and *associative areas*, which are typically involved in more complex processing tasks. The frontal lobe is involved in various sensory functions that require planned activities. However, precise location for a specific cognitive function is impossible determine, because several neural sites are processing information co-operatively (Purves et al., 2004; Martin, 2003).

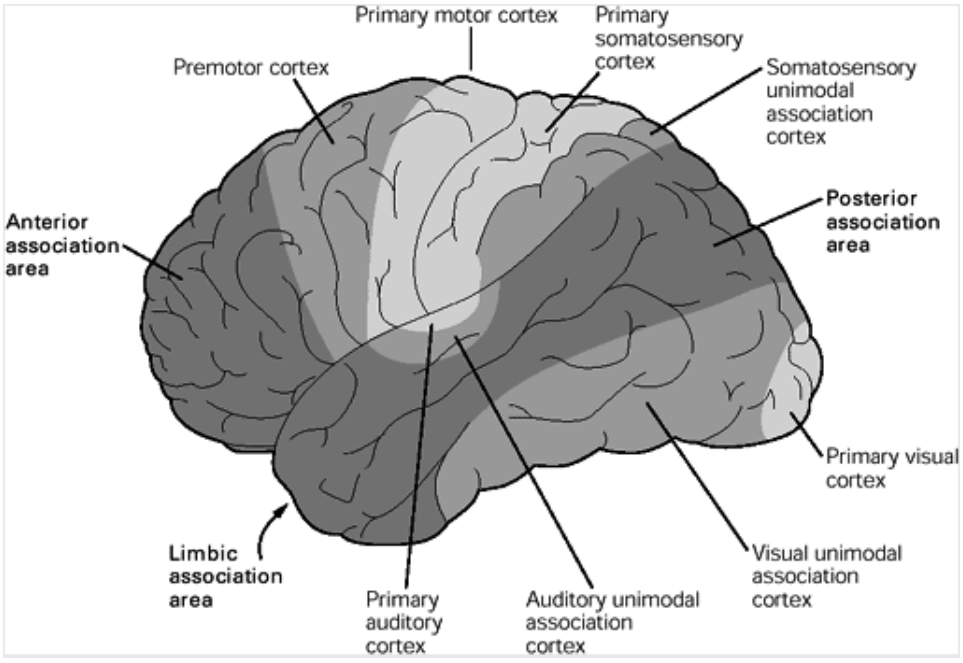



Figure 2.9: The functional division of cerebral cortex (adapted from Kandel et al., 2000).

### 2.2.2 Auditory pathway

The central auditory nervous system consist of *auditory pathway* and the *auditory cortex* (AC), which is the main processing hub of auditory information. However, much of the

neural processing occurs before the auditory inputs even reach the auditory cortex. The auditory pathway is a complex multi-phased structure of connections that route auditory information to the higher levels of the central nervous system. The left and right auditory nerves that are composed of 30,000 auditory nerve fibres each, so that main lines cross over; the left ear is innervated to right hemisphere and the right ear to left hemisphere (Seikel et al., 2009). The majority of connections are crossing over to the *contralateral* side of the brain, but the some are having bilateral connectivity. This redundancy can avoid a complete hearing loss in situations where auditory nervous system is getting damage (Martin, 2003; Purves et al., 2004).

The auditory pathway is a highly parallel system where the connections are distributed to several fiber bundles. The nerve fibers are arrayed according to the peripheral innervation of the cochlea and thus, the frequency and loudness information is stored in the compositions of neural connections (Seikel et al., 2009). The *ascending* fiber bundles terminate to the higher level *nucleus* (plural *nuclei*) which is a cluster of densely packed neuronal cell bodies. The information is also feed back to the lower level nuclei with *descending* fiber bundles. The main auditory relay nuclei (Figure 2.10, p. 14) in the brain stem are: cochlear nucleus (CN) located in the medulla, the superior olivary nuclear complex (SOC) in the pons and the inferior colliculus (IC) in the midbrain. The medial geniculate nucleus (MGN) is the thalamic auditory relay nucleus in the diencephalon and is the last processing level before the auditory cortex (Martin, 2003; Purves et al., 2004).

The first auditory relay nuclei in the auditory pathway is the cochlear nucleus. It receives information from the cochlea via the cochlear nerve[11] and diverges it to three tonotopically organized subdivisions: to anteroventral, posteroventral and dorsal part. The ventral parts have ascending connections to the superior olivary complex and the dorsal part has ascending connections to inferior colliculus (Purves et al., 2004).

The superior olivary complex is the first site in the auditory pathway that receive inputs from both ears[12] (Seikel et al., 2009). Superior olivary complex has three major components: the medial superior olivary nucleus (MSO), the lateral superior olivary nucleus (LSO), and the nucleus of the trapezoid body. High-frequency information arrives from CN to LSO and low-frequency information MSO (Seikel et al., 2009; Martin, 2003).

The inferior colliculus in the midbrain integrates all auditory information from lower auditory structures. It receive bilateral inputs from lateral superior olivary nuclei and inputs from cochlear nuclei via the lateral lemniscus. Inferior colliculus has three subdivisions: central, dorsal and external nucleus. The central nucleus is the principal auditory relay nucleus that projects to the medial geniculate nucleus (Seikel et al., 2009).

The medial geniculate nucleus is the thalamic nucleus in the diencephalon that is the last neural processing hub before the auditory cortex. The MGN has ventral, dorsal and medial subdivisions. The ventral subdivision is the principal auditory relay nucleus that has ascending connections directly to the auditory cortex (Seikel et al., 2009). The dorsal

[11] Part of Cranial nerve VIII.

[12] Processing of the interaural level difference (in LSO) and interaural time difference (in MSO) relates to the spatial audio analysis.

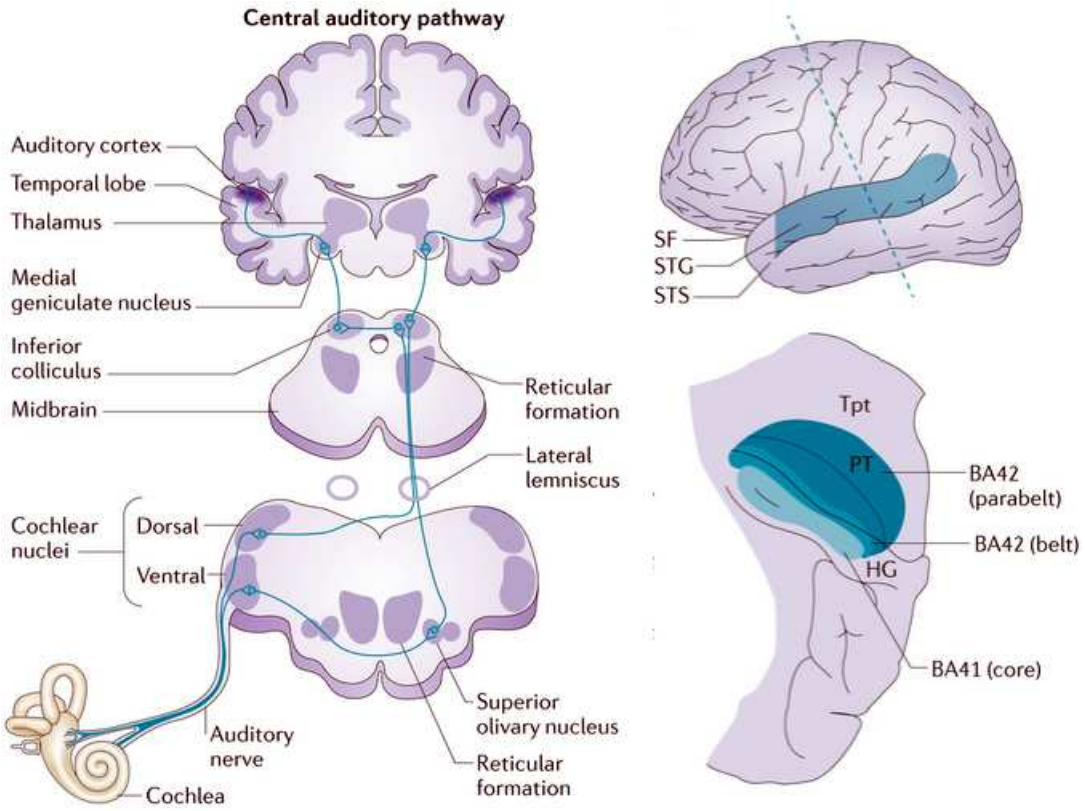



Figure 2.10: The auditory pathway (left) and the auditory cortex (right) (adapted from Javitt and Sweet, 2015).

and medial subdivisions receive information also from other sensory systems. Thus, it has been suggested that medial geniculate nucleus is involved also in cognitive control and attentional functions. For example, when attention is directed to a specific auditory feature, such as frequency (Freberg, 2009).

### 2.2.3 Auditory cortex

The auditory pathway ends to the *auditory cortex* (AC) which is the primary receptive area that represents the physiological properties of sounds by activation patterns of neurons. These complex activation patterns then reflect the emergent phenomenon – the perception of sound. However, the auditory cortex AC is not the endpoint of the human auditory system. It interacts dynamically with other cortical systems (also across the hemispheres) and even with the cochlea by the descending auditory pathway. The auditory cortex is located bihemispherically on the lateral side of the head, slightly above the pinna. The auditory information cross over in the ascending auditory pathway and therefore, the left AC get afferent inputs from the right ear and the right AC from the left ear (Baars and Gage, 2010; Seikel et al., 2009; Purves et al., 2004).

The anatomical location is on the temporal lobe where it is placed within the Sylvian fissure (SF) on the surface of the of the superior temporal gyrus (STG). These anatomical landmarks correspond to Brodmann's area 41 (BA41) and Brodmann's area 42 (BA42) (Figure 2.10, p. 14). The neural representation of auditory information is based on the function and organization of neurons in cortex. Neurons with similar receptive field properties are mingled together into functional ensembles according to stimulus *feature*; such as frequency, intensity, timing or binaural interaction. In tonotopy, neurons are organized to cortical colums based on the characteristic frequency – frequency to which the neuron is most sensitive to. These columns then form tonotopic structures of auditory cortex (Baars and Gage, 2010; Seikel et al., 2009; Purves et al., 2004).

The auditory inputs arrive first to the *primary auditory cortex*[13] (PAC), which is located in the Heschl's gyrus (HG) and corresponds the BA41 (Figure 2.10, p. 14). The primary auditory cortex is the initial reception area of the auditory information. Auditory sensations reach perception only if the auditory inputs are received and processed by the primary auditory cortex. Injuries in these areas can lead to inability to perceive sounds (Baars and Gage, 2010; Purves et al., 2004). The primary auditory cortex is tonotopically organized. However, the tonotopy is not trivial and unambiguous, but there exists multiple tonotopic neural structures in the primary auditory cortex and in the surrounding areas (Schreiner et al., 2011; Humphries et al., 2010; Talavage et al., 2004). This is logical since auditory signal is distributed and recomputed in many levels in the auditory pathway.

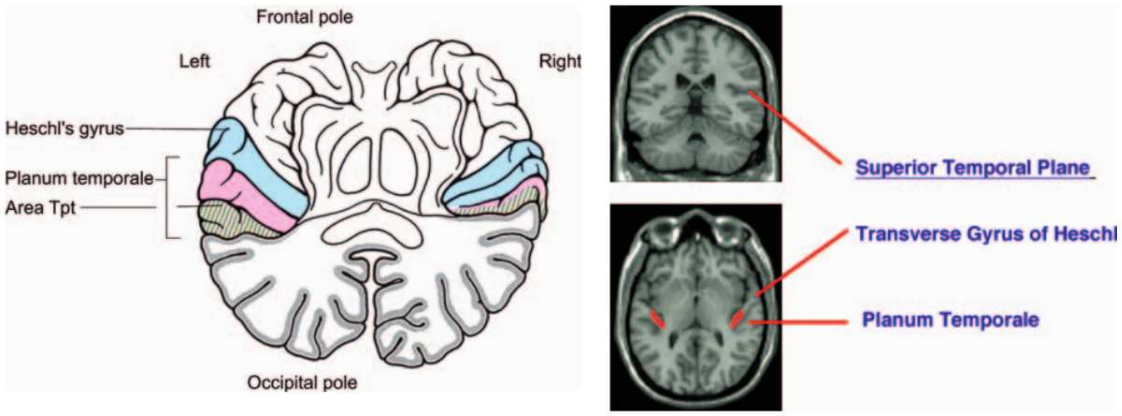



Figure 2.11: The auditory cortex asymmetry (adapted from Baars and Gage, 2010).

The information is projected from the primary AC to the surrounding *secondary auditory areas*. The first surrounding perimeter is the *belt area* (BA42) and the outermost perimeter is the *parabelt are* (Figure 2.10, p. 14). The parabelt area has connections to the temporal, frontal and parietal lobes (Baars and Gage, 2010; Purves et al., 2004). The identification of auditory objects, their properties and spatial locations requires co-operation of several cortical fields. A rough approximations of the functional hierarchy is that the primary auditory cortex extract the basic physical features of sounds (e.g. frequency,

[13] a.k.a. auditory core

intensity, duration etc.) and secondary auditory areas are involved in more complex processing. When the complexity of analysis increases, the larger areas of neural networks are involved in the processing (Seikel et al., 2009). When attention is directed to the property of a sound versus to the location of sound the information is processed differently in auditory cortex. Study of Rauschecker and Romanski (2011) discovered two separate functional streams: "what" stream for sound identification and "where" stream for sound source localization. These two streams of neural activation[14] have been discovered with both, human and animal (primates) studies (Rauschecker and Romanski, 2011).

The aforementioned physio-anatomical divisions of auditory cortices are only generalization of a typical human brains, and there exists certainly anatomical differences between humans (e.g. handedness). Moreover, the left and right auditory cortices are only roughly identical and they have differences in anatomy and physiology (Figure 2.11, p. 15). Typically the left auditory cortex is dominant in linguistic processing. The both the Wernicke's area (speech processing) and the Broca's area (written and spoken language comprehension) are located on the left auditory areas (Baars and Gage, 2010). In addition, the study of Zatorre et al. (2002) suggests that the right auditory cortex is more dominant in spectral processing, whereas left auditory cortex is more dominant in temporal processing.

## 2.3 The electrical brain activity

### 2.3.1 Electrochemical communication of neurons

The brains ability to transfer and process information is based on the dynamics of electrically and chemically active neurons. A neuron (Figure 2.12, p. 17) can receive excitatory and inhibitory impulses from thousands of other neurons with their receptive branches *dendrites*. Electrochemical stimulation that is received from other neurons conducts to the cell body called *soma*. If the excitatory stimulation is powerful enough, the *firing threshold* in soma will be exceeded and a discrete voltage spike *action potential* (AP) is generated. Action potentials travel along a wire like structure called *axon*, which allow signals to be transmitted rapidly ($\sim$ 100 m/s) within the neuron (1–20 $\mu$m). The voltage spike travels until it reaches the axon terminals that have junctions called *synapses*. This provokes the release of chemical messengers *neurotransmitters* to the synaptic cleft (gap between neurons). Neurotransmitters diffuse from *presynaptic* terminals to the *postsynaptic* terminals and bind to the receptors that are called spines. The aforementioned binding triggers *postsynaptic potential*, which is the central mechanism in cell-to-cell communication (Malmivuo and Plonsey, 1995).

The electrical excitability of neurons (APs) is based on the *cell membrane*'s ability to control ion movements across the membrane. The cell membrane is a lipid bilayer that works as a barrier between the intracellular and extracellular fluids. The both solutions

[14]The "what" stream arise from the anterior auditory cortex and projects to rostroventral prefrontal areas. The "where" stream originates from the caudal auditory cortex and projects to the caudodorsal frontal lobe (Rauschecker and Romanski, 2011)

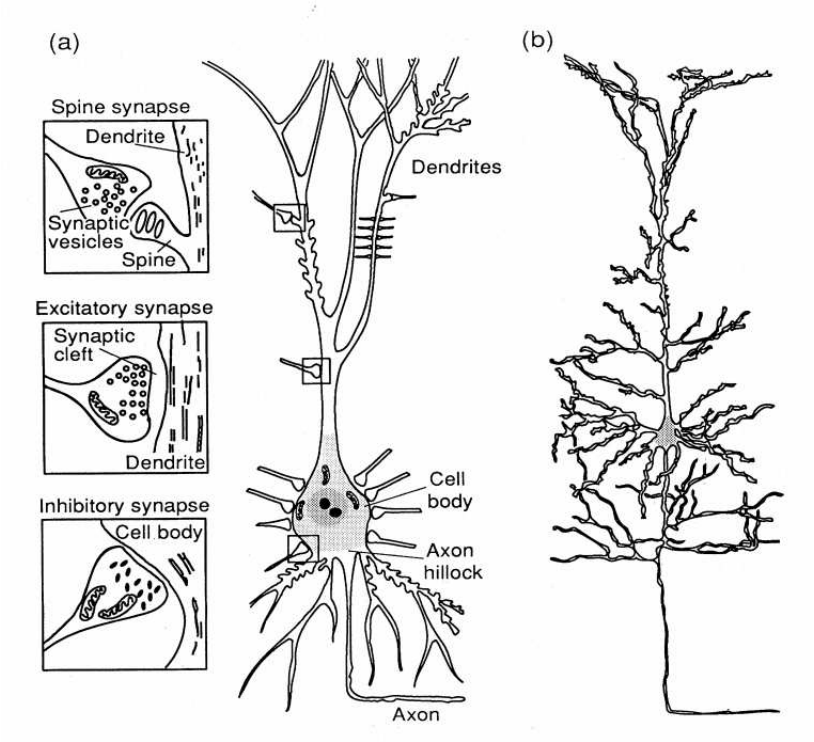



Figure 2.12: Schematic illustration of cortical neuron (a) and pyramidal neuron (b) (adapted from Hämäläinen et al., 1993).

have mobile ions: sodium ($Na^+$), potassium ($K^+$), chloride ($Cl^-$), and calcium ($Ca^{2+}$). The cell membrane has also embedded protein structures that act as semipermeable *ion channels*, or as metabolically driven *ion pumps* that push ions through these channels. The ion pumps are "charging" the cell by pushing more *cations* to the extracellular fluid, which leads to a potential difference across the cell membrane. This *membrane potential* is about -70 mV at the resting state. The voltage and concentration differencences cause *electrochemical gradient* that tries to drive ions to move across the membrane towards the electrochemical equilibrium state. Thus, the influx or outflux of ions through a specific ion channel depends on the electrochemical gradient (Malmivuo and Plonsey, 1995).

The function of active cell membrane can be modeled with electrical circuit presented in figure (Figure 2.13, p. 18) where the ionic currents through the resistances are divided in sodium, potassium and *leakage components*, which are in parallel with capacity. This is known as "Hodgkin and Huxley model of cell membrane" (Hodgkin and Huxley, 1952).

Action potentials are all-or-nothing type of events that are triggered if the firing threshold exceeds as a result of excitatory stimulation. The postsynaptic side of a nerve cell receives thousands of presynaptic inputs from other neurons, which cause transient changes in the membrane potential. The excitatory stimulation opens the *chemically gated ion channels* and positively charged ions flow into the postsynaptic cell. This *excitatory postsynaptic potential* (EPSP) *depolarizes* the membrane potential, since the inflow of cations make the interior voltage less negative. Correspondingly, inhibitory stimulation cause *inhibitory postsynaptic potentials* (IPSPs) where cations flow out from cell through ion channels,

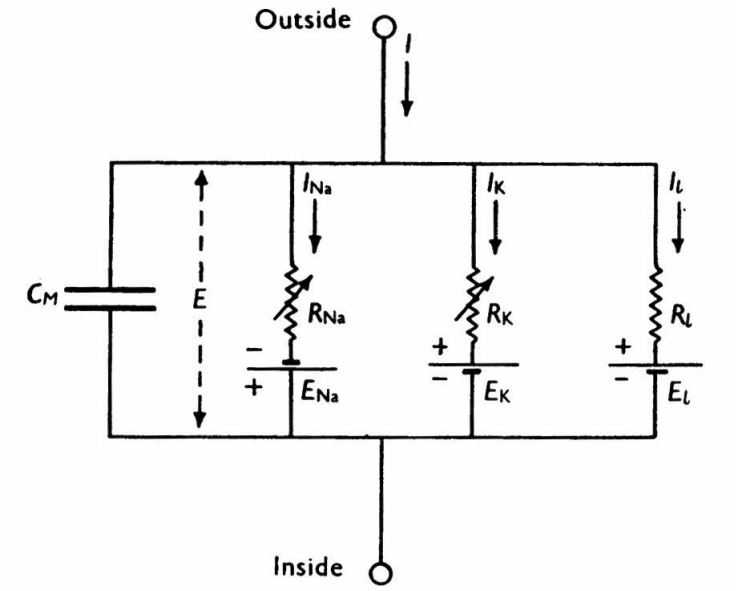



Figure 2.13: Hodgkin and Huxley model of cell membrane (adapted from Hodgkin and Huxley, 1952).

which makes the interior voltage more negative (*hyperpolarization*). These weak postsynaptic currents flow along dendrite all the way to the *axon hillock* by slowly decaying because of the resistance of the intracellular fluid. If the postsynaptic currents depolarize the membrane potential at axon hillock enough, the *voltage gated ion channels* open and cations (sodium) flow inside the cell. The depolarization triggers a chain reaction where all adjacent sodium channels to open one by one along the axon. The depolarization is followed by a similar chain reaction where outflux of potassium ions cause a hyperpolarization. Thus, the membrane potential goes through a rapid voltage swing that travels like a pulse along the axon until it reaches the axon terminals. The initial condition is restored after the action potential by the *sodium-potassium pumps* that transport sodium out and potassium in until the cell is in again in resting potential (Malmivuo and Plonsey, 1995; Luck, 2005).

### 2.3.2 The generation of extracranial electromagnetic fields

The bioelectric events relating to neural communication produce weak electric fields that can be detected with suitable instruments (e.g. by placing microelectrodes into the cortical tissue). Sometimes these microscopic currents can sum up and generate macroscopic electromagnetic fields that can be detected from a distance, even from the surface of the scalp. Typically the macroscopic fields are generated in the cortical layers where large neurons, *pyramidal cells* are aligned parallelly to cortical columns, perpendicular to the cortex plane (Hämäläinen et al., 1993). The aforementioned tiny current sources are spatially and temporally coherent, and thus, synchrononous mass activity can give raise to larger fields. The superposition principle states that in linear systems, the net current at a given place and time is the algebraic sum of microscopic currents (Ramírez et al., 2010).

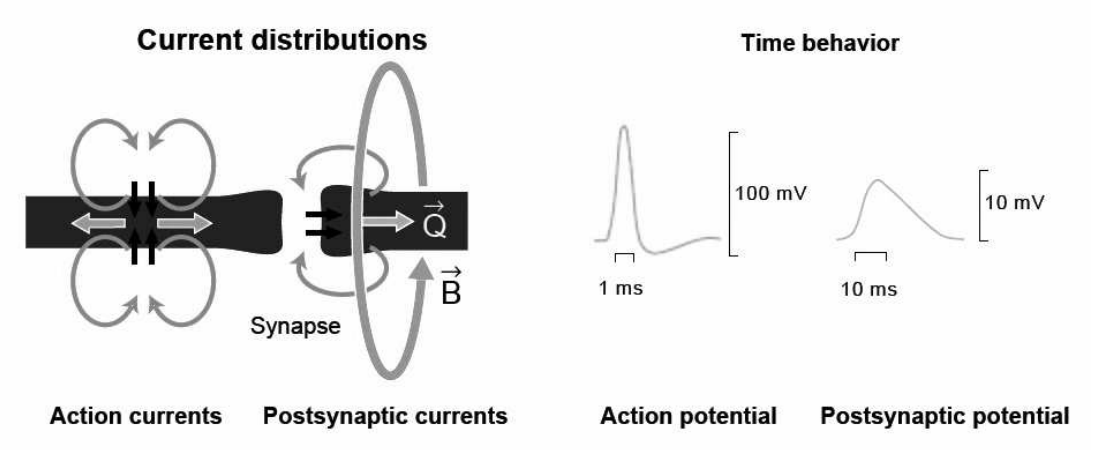



Figure 2.14: A schematic illustration of action potential and postsynaptic potential (left) and the latencies (right) [adapted from lecture slides of Hämäläinen].

The potential microscopic current sources (AP and PSP) that can generate macroscopic fields are presented Figure 2.14 (p. 19). The current distribution of the postsynaptic potential has two opposing charges in close vicinity that from a current dipole ($\mathbf{Q}$) along the dentrite. In the equation,

$$\mathbf{Q} = I\lambda \tag{2.2}$$

$\mathbf{Q}$ is the current dipole vector, $I$ is the intracellural current and $\lambda$ is the length constant of the exponential decay[15], which is typically 0.1 – 0.2 mm in cortical neurons (Hämäläinen et al., 1993).

The current distribution of the action potential has two opposing depolarizing and repolarizing current sources that form a quadrupole (Figure 2.14, p. 19). The electromagnetic field of quadrupole is remotely zero, because two reverse dipoles cancel each other out. Moreover, action potentials are transient events ($\sim$ 1 ms), which makes the temporal summation less likely. Instead, the temporal summation of PSPs is possible, because the duration is from 10 ms to hundreds of ms. In addition, a dipolar fields that are generated by PSP decay with distance of $1/r^2$, whereas the quadrupolar field generated by APs decrease with distance of $1/r^3$ (Hämäläinen et al., 1993).

Thus, the postsynaptic potentials with dipolar current distribution and appropriate timescale are the principal generators of electrical fields that sum up and polarize the cortical tissue widely. The electric current associated with specific neuronal activity (e.g. PSPs) cause a *primary current* $\mathbf{J_p}$ that travel in the *conducting medium* (brain) as a secondary *volume current* $\mathbf{J_v}$. The cortical tissue can be modeled as a volume conductor, where the current conducts to all three dimensions from the source. The total current density can be expressed with equation,

[15]The lenght constant is dependent on the conductance of the membrane and on the resistance of the intracellular fluid per unit length ($\lambda = (g_m r_s)^{-1/2}$).

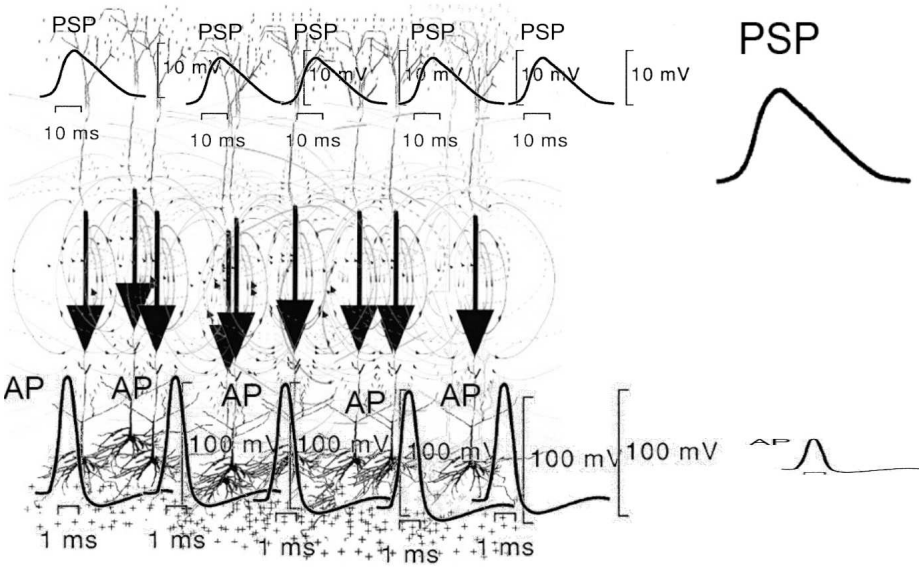



Figure 2.15: Mass effect of neural ensembles [adapted from lecture slides of Hämäläinen].

$$\mathbf{J_{tot}} = \mathbf{J_p} + \mathbf{J_v} = \mathbf{J_p} + \sigma\mathbf{E} \tag{2.3}$$

where $\sigma$ is conductivity for infinite and homogeneous volume conductor (Ramírez et al., 2010; Singh, 2006).

The majority of these macroscopic fields originate in the cortical layers where large neurons, *pyramidal cells* are organized parallelly to cortical columns. These apical dendrites of pyramidal cells are aligned perpendicular to the cortex plane. Excitatory PSP causes a net negativity (sink) at the apical side of dendrite (Figure 2.15, p. 20) and a net positivity (source) at the side of the cell body. The mass activity of EPSPs in apical dentrites give rise to an *open field* current distribution, which polarize the brain tissue also from a distance. Instead, the neural mass activity of IPSPs form a *closed field* current distribution whose net polarization zero from a distance, since the dentrites are organized radially around the soma (da Silva, 2004; Hämäläinen and Hari, 2002).

If we replace the term $I$ with $\Delta V/\lambda$ in Equation 2.2, the current dipole can be expressed in form:

$$\mathbf{Q} = \sigma_{in} A \Delta V \tag{2.4}$$

where $\sigma_{in}$ is the intracellural conductivity, A the crossectional area of a dentrite and $\Delta V$ the potential change caused by the PSP. With typical values $\sigma_{in} = 1\ \Omega^{-1}\ m^{-1}$, $A = \pi/4 \cdot 10^{-12}\ m^2$ (where $d = 1\ \mu m$) and $\Delta V = 25\ mV$ the equation gives $\approx 20\ fAm$ for an elementary postsynaptic current (Hämäläinen et al., 1993). The current densities that can generate measurable extracranial EMFs are typically between 10 to 20 nAm. Thus, approximately a million active synapses from 50 000 pyramidal cells are needed. This

corresponds $\approx 1\ mm^2$ area at cortex (Hämäläinen and Hari, 2002). However, the area should be larger ($\approx 25\ mm^2$) because currents flow also in opposite directions, which diminish the overall effect (Ramírez et al., 2010; Singh, 2006).

According to Amperes law, all moving charged particles (ions) produce magnetic fields. The current flow along a wire-like conductor (e.g. dentrite) generates a magnetic field around it, so that the magnetic field lines form concentric circles around wires cross-section. The direction of magnetic field lines can be determined by using the "right hand grip rule". Thus, the bioelectric fields that are generated mainly by synchronous mass activity (EPSP) of pyramidal neurons are coupled to the *biomagnetic fields*. According to *Maxwell's equations* electric fields are coupled to the magnetic fields as follows (Hämäläinen et al., 1993):

$$\nabla \cdot \mathbf{E} = \frac{\rho}{\epsilon_0} \tag{2.5}$$

$$\nabla \times \mathbf{E} = -\frac{\partial \mathbf{B}}{\partial t} \tag{2.6}$$

$$\nabla \cdot \mathbf{B} = 0 \tag{2.7}$$

$$\nabla \times \mathbf{B} = \mu_0(\mathbf{J} + \epsilon_0 \frac{\partial \mathbf{E}}{\partial t}) \tag{2.8}$$

where $\mathbf{E}$ is the electrical field density [V/m], $\rho$ is the volume charge density [$C/m^3$], $\epsilon_0$ is the electrical permittivity of free space [F/m], $\mathbf{B}$ is the magnetic flux density [T], $\mu_0$ is the magnetic permeability of free space [H/m] and $\mathbf{J}$ is the total current density [$A/m^2$]. Because the frequency range of neural signals is less than 100 Hz, the *quasistatic approximation* can be applied in Equation 2.6 and Equation 2.8 to omit the terms $\partial\mathbf{B}/\partial t$ and $\partial\mathbf{E}/\partial t$ (Hämäläinen et al., 1993).

### 2.3.3 Non-invasive measurement methods

As was stated in the previous chapter, coherent neural mass activity can generate extracranial signals. The amplitudes of these signals are high enough to be monitored with non-invasive neuroimaging methods, such as *electroencephalography* (EEG) and *magnetoencephalography* (MEG). In both technologies a single sensor detects EM field changes at a constricted cortical area. The single channel signal reflects the temporal dynamics of the primary and secondary currents (a.k.a. volume current) that are close to the sensor. The modern measurement devices are *multi-channel systems* whose sensor arrays are designed to cover the whole cortex. The spatial information is based on the sensor locations. However, the intracranial neural sources cannot be located unambiguously by analyzing the sensor data of the extracranial field. These *inverse problems* are always *ill-posed*, which means that one or several intracranial source combinations can generate identical

activation pattern at the sensor side. However, the number of potential sources can be reduced by *magnetic resonance imaging* (MRI) technology that captures structural images of individual brain anatomy. The high spatial accuracy of MRI images can be combined with the high temporal accuracy of EEG and MEG data (MRI constrained MEG). The most trivial solution to the inverse problem is to approximate the actual neural source with the idealized current source – *equivalent current dipole* (ECD) (Luck, 2005).

The other non-invasive neuroimaging technologies: *functional magnetic resonance imaging* (fMRI) and positron emission tomography (PET) are common in studies where cognitive functions are mapped to the neuroanatomical structures. The fMRI device measures the hemodynamic response, which describes how oxygen concentration changes in blood as a function of time. High oxygen level represents increased neural activity since brain cells consume more energy under the cognitive load (Matthews and Jezzard, 2004). The PET device measures the drift of the radioactive tracer in the CNS. However, only EEG and MEG have adequate temporal accuracy (ms time-scale) to measure the transient cortical responses. Thus, it is common to use combined imaging technologies in neurophysiological research (e.g. fMRI and EEG) (Luck, 2005).

In EEG technology the electrical brain activity is measured with skin-attachable electrodes from multiple locations (typically 64 or 128 channels) on the surface of the head. The potential difference of each electrode is defined according to a stable and well-known reference electrode. The EEG signal represents mainly the volume currents arriving to the scalp from the primary sites. The EEG is most sensitive when the neural sources (pyramidal cells) are superficial current sinks that are oriented radially to the surface of the head (Figure 2.17, p. 23). The aforementioned sources show up as negative electrical field potentials in the signal and the amplitudes are typically in $\mu V$ scale. The localization of the primary sources from the EEG signal is rather difficult, because of the high resistance of the skull that attenuates and distorts the signal (Luck, 2005).

### 2.3.4 Magnetoencephalography

In magnetoencephalography (MEG) the magnetic field changes relating to electrical brain activity are measured over the head with extremely sensitive detectors called *magnetometers*. The modern MEG equipments are multichannel systems where the sensor array is designed to cover the whole cortical area. Magnetometers are sensitive to the tangential current sources (primary currents) that exist on the walls of cortical sulci (Figure 2.17, p. 23). These tangential currents generate magnetic fields around the axis of the current flow vector ($\mathbf{Q}$)[16], which spread outside the head since the skull is transparent to the magnetic fields. Thus, MEG provides better spatial accuracy of the primary current sources than EEG. The EEG signal reflect volume currents, which are distorted by the skull. However, as a disadvantage, MEG is insensitive to magnetic fields that are generated by radial currents. These fields are typically left within the head that is a spherical volume conductor (Singh, 2006; Luck, 2005).

[16]The orientation of the magnetic field lines depends on the direction of the current and be determined by the "right hand grip rule".

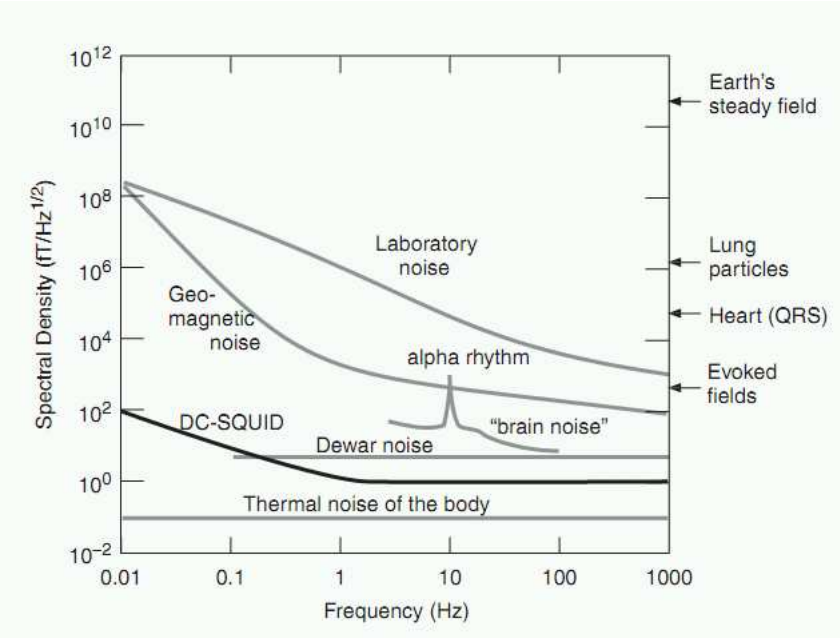



Figure 2.16: Magnetic field strengths of various sources (adapted from Hämäläinen et al., 1993).

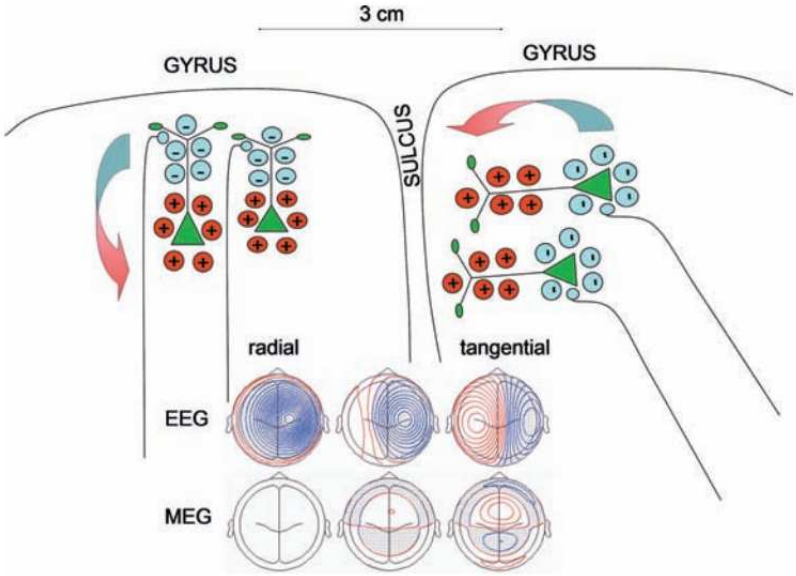



Figure 2.17: EEG and MEG sensitivity (adapted from Michel et al., 2009).

The MEG measurements are technically challenging because the cortical magnetic fields are extremely weak and the interfering fields are strong (Figure 2.16, p. 23). The signals of interest are typically between 50–150 $fT$ ($10^{-15}$ $T$) whereas the magnetic field of the Earth is almost $10^9$ times stronger, between 25–65 $\mu T$ ($10^{-6}$ $T$). Thus, the measurements should be made in a *magnetically shielded room* (MSR) that provides passive shielding against the ambient magnetic noise. The magnetically shielded room is a heavy mul-

tilayered structure, which is comprised of two or three concentric shells[17] of mu-metal and aluminum. The magnetically shielded room do not give protection against interference signals that are generated inside the room. Thus, the detection of signals of interest requires exceptional sensor technology, noise cancelation techniques and sophisticated computation (Parkkonen, 2010; Hämäläinen et al., 1993).

The sensor technology is based on a quantum mechanical phenomenon called *superconductivity*. Superconductive materials lose their electrical resistance when they are cooled below their critical temperature $T_c$, which is typically less than 20 Kelvin. The electric current flows in the superconducting loop infinitely as long as the low temperature is maintained. In the modern MEG tehchnology the detectors: *superconducting quantum interference devices* (SQUIDs) are cooled down to the superconducting state by liquid helium whose temperature is 4.2 K (-269°C). The liquid helium and SQUIDs are kept in a helmet-shaped container (*dewar*). The SQUIDs are arranged within the dewar so that the distance between sensors and head surface is as small as possible (Figure 2.18, p. 24) (Parkkonen, 2010).

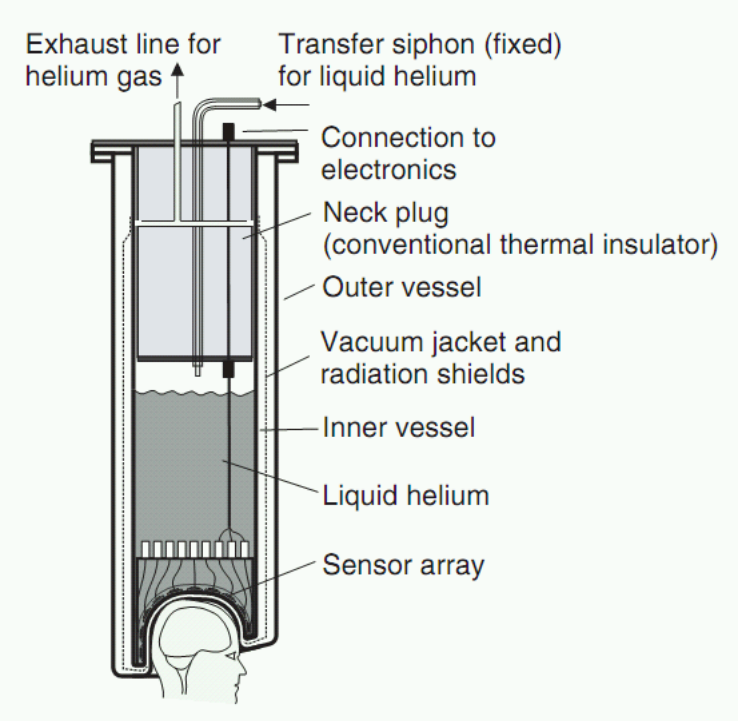



Figure 2.18: Schematic cross section of Sensor Arrays and a Dewar (adapted from Parkkonen, 2010).

When the superconducting loop is placed in a time-varying magnetic flux, a shielding current is induced around the circuit. This voltage provides an indirect measure of the magnetic field change. However, a normal current measurement system can not been connected to the loop without spoiling the measurement. The measurement can be carried

[17]Mu-metal is an alloy consisting of nickel and iron that supply protection against low frequency magnetic noise whereas aluminum provides protection against high frequency magnetic noise.

out properly by using two superconductor loops, which are coupled together by a thin layer of an electrical insulator called *Josephson junction* (Figure 2.19, p. 25). The current that crosses the juncion, the Josephson current can be measured (Hämäläinen et al., 1993).

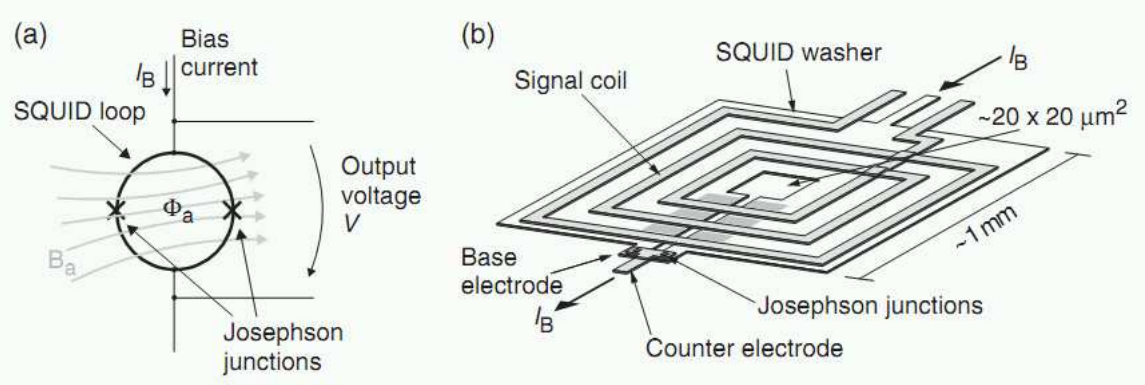



Figure 2.19: A schematic illustration of the SQUID loop (a) and a modern thin film dc-SQUID (b) (adapted from Hämäläinen et al., 1993).

However, SQUIDs itself are not suitable to couple properly to the constanly changing magnetic fields since the surface area of the loop is too small (d $\sim$ 1 mm) and its angle is rarely perpendicular to the orientation of the magnetic flux. These difficulties can be overcome by applying a *flux transformer* that collect magnetic flux from larger area and from appropriate angle. Typically the spatial derivative of the field (*gradient*) is measured in three dimensions, along x, y and z components. The gradient [T/m] describes the change rate of the field along the component vector (Hämäläinen et al., 1993).

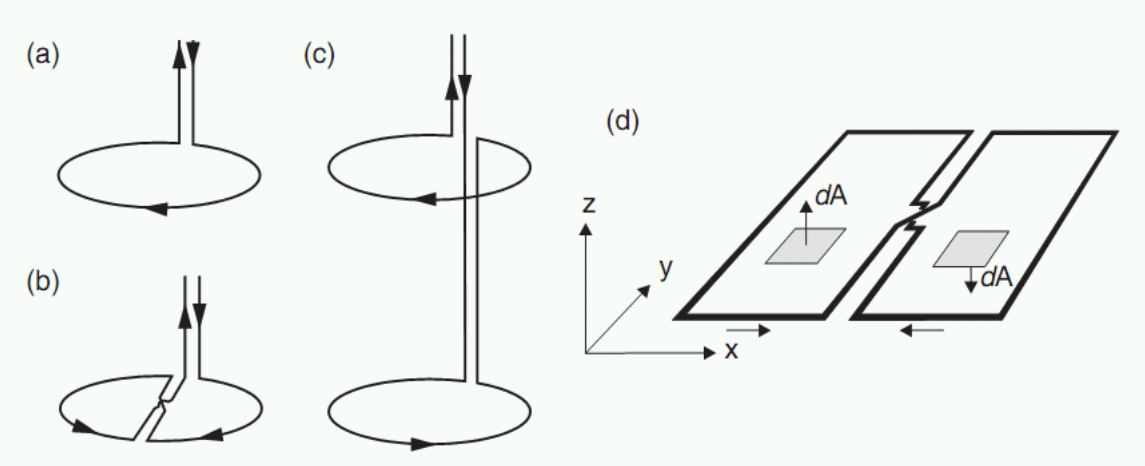



Figure 2.20: Flux transformer geometries: (a) magnetometer (b) planar gradiometer (c) axial gradiometer (d) dBz/dx (adapted from Hämäläinen et al., 1993).

The simplest flux transformer configuration the magnetometer (Figure 2.20a, p. 25) is a single pick-up coil that detects $B_z$ component (a.k.a. $B_r$ component) that is perpendicular to the cortex surface. Magnetometers are somewhat unpractical because of the sensitivity to distant sources, which are typically related to interferening fields. More practical

method is to use two-coil configuration where the signal coil is closer to the cortical source and the compensation coil further away from the source. The interference fields, which are typically from distant sources can be cancelled out by wounding the coils in opposite directions, since distant signals are roughly equal in both coils. The aforementioned coil configuration measure the spatial derivative instead of the homogeneous magnetic field itself and is known as *gradiometer*.

The spatial sensitivity patterns (*lead fields*) of different flux transformer configurations are presented in Figure 2.21 (p. 26). The *axial gradiometer* configuration detects the $\partial B_z/\partial z$ component because the coils are placed along the radial z-axis. In the *planar gradiometer* configuration the coils are placed side-by-side in the same plane to detect either $\partial B_z/\partial x$ or $\partial B_z/\partial y$ component. The planar gradiometers are most sensitive to sources that are right beneath them and the axial gradiometers to sources that are little bit deeper (Parkkonen, 2010).

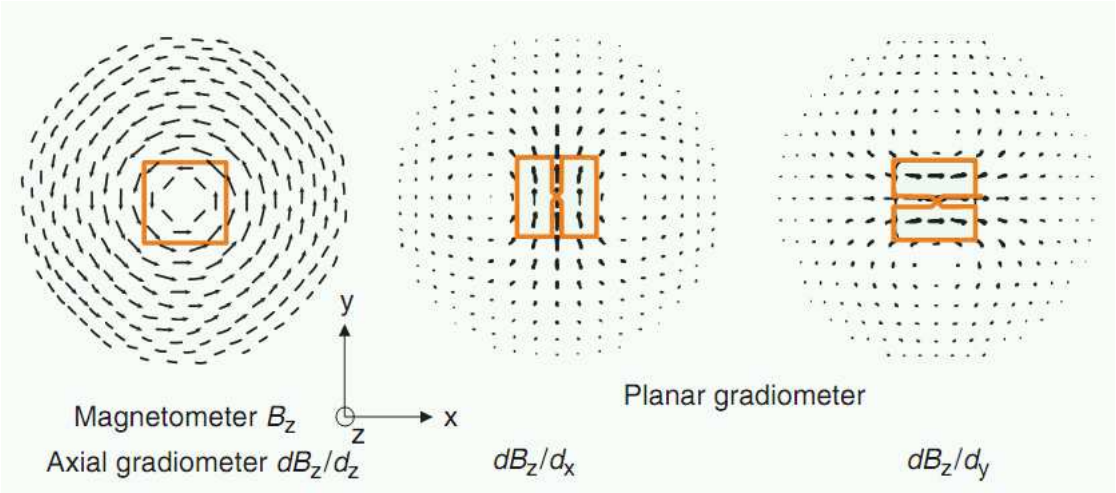



Figure 2.21: The lead fields of different coil configurations (adapted from Parkkonen, 2010).

The modern MEG systems have typically triplets where two orthogonally oriented planar gradiometers and one magnetometer are integrated in single units. The planar gradiometers measure both strength and orientation of the local source. Since the gradient orientations vary frequently during the measurements, it can be more practical to take vector sum of the gradiometer channel pairs. The underlying equation:

$$G_p = \sqrt{\frac{\partial {B_z}^2}{\partial x} + \frac{\partial {B_z}^2}{\partial y}} \tag{2.9}$$

gives information of the strength of the neural activity [18].

[18] The $\partial B_z/\partial x$ and $\partial B_z/\partial y$ are the two orthogonal field components.

### 2.3.5 The waveforms of MEG and EEG signals

The multichannel MEG and EEG systems give spatio-temporal information of the activated cortical structures. A single channel signal (*encephalogram*) reflect the ongoing population level activity at a constricted area. Stable rhythmic fluctuations on the waveform represent spontaneous brain activity, whereas transient spikes on the waveform are responses to sudden sensory events (Michel et al., 2009).

The most pronounced patterns in the encephalogram, the neural oscillations reflect global brain functions like sleep, wake or other states of consciousness. The rhythm of the neural oscillation varies according to the state of arousal. The alpha rhythms show up during relaxed wakefulness and alternates between 8 Hz and 13 Hz. The beta (13–30 Hz) and gamma (30–70 Hz) rhythms are associated with increased levels of alertness, and these rhythms occur also during rapid eye movement (REM) sleep. The delta rhythms are slow oscillations (0.5–4 Hz) that occur during unconscious deep sleep. The theta rhythms are related to conscious waking states at frequencies between 4 Hz and 8 Hz, and can take place e.g. during working memory tasks. The sigma rhythms (a.k.a. sleep cycles) occur during sleep and consist of periodically alternating oscillations (Michel et al., 2009).

### 2.3.6 Evoked responses

The high temporal resolution of EEG and MEG ($\sim$ 1–10 ms) gives an opportunity to study the information exchange between the cerebral cortex and the environment (for a review see e.g. (Kotchoubey, 2006)). In the *evoked response paradigm* the sensory functions can be studied indirectly by observing the stimulus-related patterns in the encephalogram. The patterns that are evoked by sensory stimuli are typically transient peaks or troughs in the waveform, and are known as *event-related responses* (ERRs) or *evoked responses*. Depending on the measurement technology (EEG or MEG) the evoked responses are identified as *event-related potentials* (ERPs) or *event-related fields* (ERFs) (Zani et al., 2003).

The timing of the peak gives valuable information of the processing phase at CNS (e.g. the dependency of peak's latency and the corrsponding neural structure at CANS can be seen in Figure 2.23, p. 29). The amplitude of the peak represents the strenght of neural activation. The evoked response paradigm can be applied in the research of different sensory functions, especially in the auditory, visual and somatosensory domains. The corresponding responses are termed as *auditory evoked responses* (AERs), *visual evoked responses* (VERs) and *somatosensory evoked responses* (SSERs) (Zani et al., 2003).

The amplitudes of evoked responses are a few decades smaller than the amplitudes of the spontaneous brain activity (Figure 2.16, p. 23). Therefore, the non-relevant fluctuations and other interfering signals should be filtered out by averaging. The filtering can be made by measuring short segments of encephalogram data (*trials*) that are time-locked to the fixed stimuli. The desired event-related response can be fetched out by calculating the average of the trials (Figure 2.22, p. 28). Typical 100 trials per evoked response will

ensure good signal-to-noise for the ERR signal (Parkkonen, 2010).

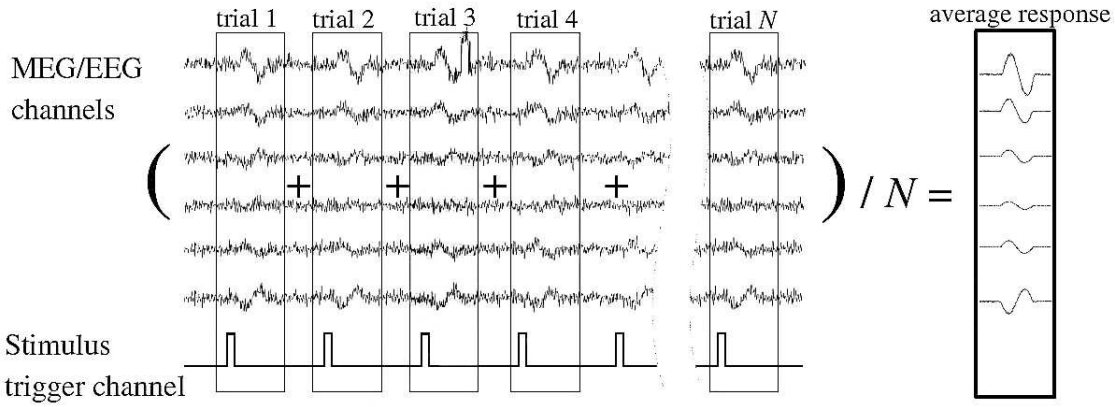



Figure 2.22: Stimulus-locked time-domain averaging [adapted from lecture slides of Parkkonen, 2005].

The evoked responses are functions that represent event-related neural activity with a series negative and positive amplitude deflections with respect to the *baseline*. The baseline is a hypothetic zero line that represent the nervous system state during normal spontaneous neural activity ("idle state"). The baseline related amplitude gives information of the strength of the event related neural activity. However, sometimes it can be more practical to use *peak-to-peak amplitude*[19] to describe the signal strength – for example, if there is drift in the baseline (Parkkonen, 2010).

The latency of a peak or a trough in the ERR-waveform gives information of the stage of the processing in the sensory system. In wider perspective the latency gives also knowledge of information exchange between the organism and the environment, since a stimulus provokes also a chain of behavioral events: perception, cognition and action. All these behavioral events have also neural correlates that can be studied by event-related response paradigm (Kotchoubey, 2006).

A pattern in the ERR-waveform (e.g. a peak) is defined as a *component* if it describes a specific sensory or cognitive function. The components are named according to the *polarity* (P or N) and latency (rounded numeral in ms) of the peak value. For instance, a component with a negative polarity that appears about 100 ms after the stimulus onset is defined as N100m. The attribute m specifies that the measured component is a magnetic flux density. However, the positive and negative attributes are somewhat misleading, especially if the measurements are made with MEG (Zani et al., 2003; Hämäläinen et al., 1993).

[19]The difference between positive and negative maximum.

### 2.3.7 Auditory evoked responses

Auditory evoked responses (AERs) or *tone-evoked responses* are stimulus-related patterns in the ERR-waveform that are triggered by auditory stimuli. The *auditory evoked fields* (AEFs) are measured with MEG and *auditory evoked potentials* (AEPs) with EEG. The latencies of different AER components depend on the stage of processing in the central auditory nervous system (Figure 2.23, p. 29). The first components, the *auditory brainstem responses* (ABRs) are expressions of the auditory processing, which originate from the initial stages of the auditory pathway[20]. Most of the signals originate from the auditory brainstem. ABRs show up 1 ms – 12 ms after the stimulus onset and are indexed by Roman numerals (I–VI) (Zani et al., 2003; Luck, 2005).

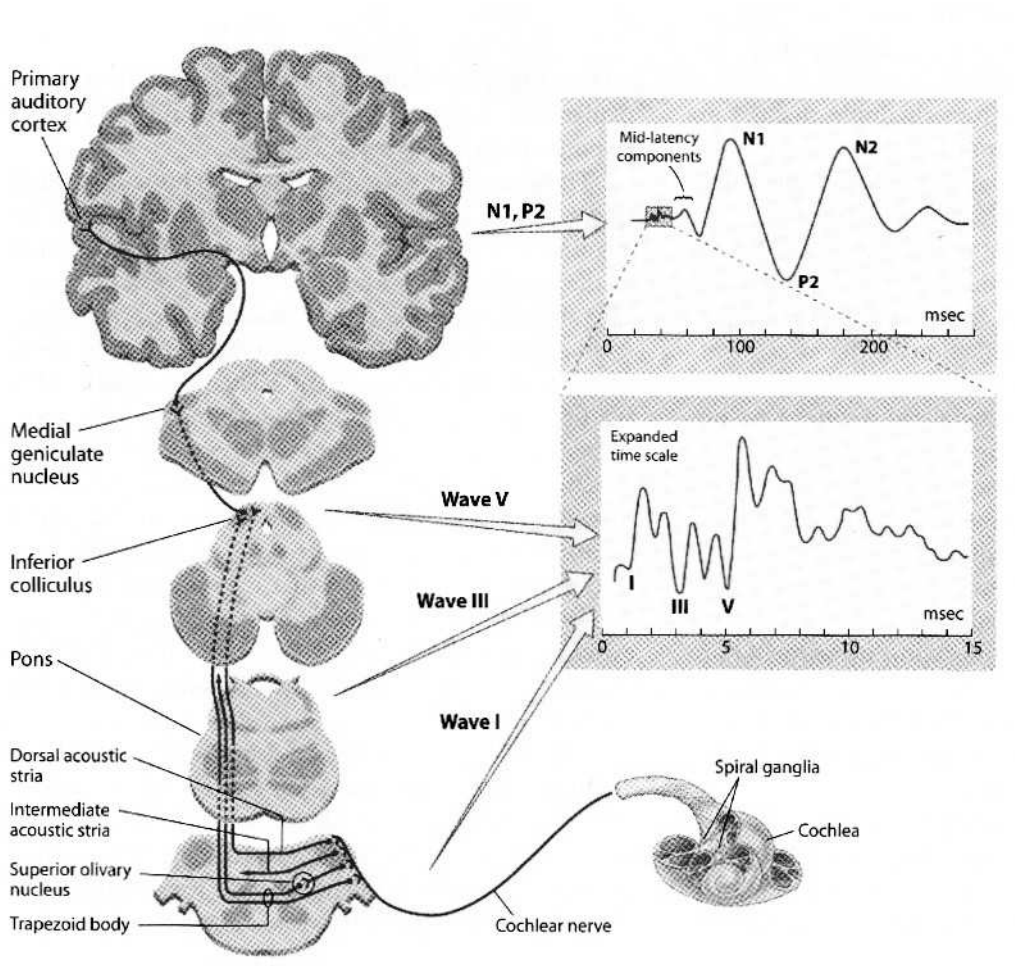



Figure 2.23: Event-related responses and the possible locations of sources in the human auditory system (adapted from Gazzaniga et al., 2002).

The ABRs are followed by *middle-latency responses* (MLR), four peaks (P0, Na, Pa and

[20] Auditory nerve, the cochlear nuclei, the superior olives, the lateral lemniscus and the inferior colliculus (Zani et al., 2003)

Nb) in the waveform that appear between 12 ms and 50 ms from the stimulus onset. The middle latency responses are expressions of the auditory processing that originate from inferior colliculus, medial geniculate nucleus and possibly from the initial stages of the primary auditory cortex (Zani et al., 2003; Luck, 2005). The *long-latency responses* (LLR) (P50, N100 and P200) are group of cortical components whose latencies are approximately 50 ms, 100 ms and 200 ms from the stimulus onset. The middle-latency and long-latency responses reflect typically the processing of physical sound characteristics, such as frequency, intensity, duration and onset of the auditory stimulus. However, the aforementioned components are also sensitive to cognitive functions, such as attention (Zani et al., 2003).

The most prominent peak in the ERR-waveform – N100(m) component – is also known as *onset response* since it is triggered by stimulus onset[21] or a sudden change in a continuous sounds (e.g. change in sound intensity or frequency). Thus, the component is commonly used in perceptual studies where neural responses are compared to the temporal changes in the auditory environment. The prominence of N100(m) component results from the fact that majority of cortical neurons respond to the stimulus onset, but not to the prolonged stimulation (Näätänen and Picton, 1987). The sources of N100 components have been localized to the primary (Hari et al., 1980) and secondary auditory cortices (Pantev et al., 1995), as well as to the belt and parabelt areas (Inui et al., 2006). Anatomically these areas correspond to Heschl's gyrus, superior temporal gyrus, and the planum temporale (Zani et al., 2003).

Since N100(m) is an onset response, the component amplitude is also dependent on the interval time between the preceeding stimuli, the *interstimulus interval* (ISI). If the interstimulus time is shortened, the N100 peaks are suppressed in amplitude (Rosburg et al., 2004). The suppression results from *synaptic depression*, which means that the refractory period that cell assemblies need to recharging is too short (Budd et al., 1998).

In contrast to the transient responses, the evoked component could be also a "stretched" peak that typically appears when the triggering stimulus is temporally long enough. *Sustained response* is a baseline shift (electric or magnetic) that holds up to the end of the stimulus. *Sustained potential* (SP) is the term for a DC-shift in EEG measurements and *sustained field* (SF) the term for the magnetic counterpart (Picton et al., 1978; Pantev et al., 1994). Sustained responses are typically less sensitive to the manipulation of the stimuli, which indicate that the neurophysiology differs from the transient responses (Picton et al., 1978; Pantev et al., 1994). The sources of sustained responses have been localized to the primary auditory cortex (Hari et al., 1980; Pantev et al., 1994) as well as to the belt and parabelt areas (Okamoto et al., 2011).

[21] The rise time of the sound onset should be short enough to provoke N100(m) response (Zani et al., 2003).

## 2.4 Psychophysical method

In the previous sections were reviewed, how to measure physiological data of the neural functions with non-invasive methods. However, the neural activation patterns at the auditory cortex relate (or translate) ultimately to the subjective auditory experiences. Thus, the sensory functions can be studied also by the methods of experimental psychology – since the physical stimuli provoke also mental responses: sensations and perceptions. According to the *cognitive theory* the functions of the mind can be studied quantitatively with scientific methods (Niiniluoto and Koskinen, 1995).

In *psychophysics* the sensory and perceptual functions are investigated indirectly by observing the relation between the stimulus and the sensation it provokes. The psychophysical method exploits the *black box approach* where larger functional entities can be studied by analyzing the behavior of the input-output (stimulus-response) pairs. Thus, the inner structure of a complex system (e.g. the auditory pathway) can be partially ignored. Naturally, also the physiological measurements and the knowledge of system's inner structure can be exploited in the psychophysical studies. Psychophysics combine psychological and physiological research methods that can be applied in the research of sensory functions, especially in the auditory, visual and somatosensory domains (Green and Swets, 1988).

In a typical psychophysical experiment subject's experience or behaviour is observed and measured with respect to a systematically changed stimulus feature(s). The psychophysical experiments are based on population level data where a group of subjects are making tests with identical stimulus setup. The stimulus setup is usually kept trivial so that the stimuli can be measured and controlled objectively in the experiment. The psychophysical measurements are scientifically valid if the experimental framework is convergent with the real-life-phenomenon and the biasing factors are eliminated with proper statistical methods. The test procedures should be well documented so that the findings can be reproduced and validated (Green and Swets, 1988).

The experimental data is used typically to construct *psycho-metric functions*, which are quantitative assessments or visualizations of the dependency between objective and subjective data. The dependency can be, for instance, how the magnitude of the physical stimulus corresponds subject's experience of the stimulus magnitude. One of the most typical psycho-metric function is the *threshold* (a.k.a. *sensory limen*), which can be used to describe the dynamic range, limits or accuracy of the sensory system. The sensory thresholds are quantified mainly with performance measurement experiments (Green and Swets, 1988).

The *absolute threshold* (a.k.a. *detection thresholds*) is the minimum intensity level for a stimulus at which the subject is able to detect it. Typically 50% p-level is used in the determination of the absolute threshold, which means that every second sample is detected with the particular intensity level. The *discrimination threshold* is the smallest intensity difference between two stimuli that can be perceived, and is called as *just-noticeable difference* (JND). Psychophysical experiments have showed that the just-noticeable difference is a

constant proportion $k$ despite variation in magnitude. The $k$ can be approximated with the underlying equation:

$$\frac{\Delta S}{S} = k \tag{2.10}$$

where ($S$) is the signal magnitude and $\Delta S$ is the magnitude difference. This *Weber's constant* holds true for most sensory domains as long as the stimulus intensity is not too close to the sensory threshold (Green and Swets, 1988).

The modern psychophysical experiments are based heavily on the aforementioned *threshold measurement paradigm*, but also on *signal detection theory* (SDT), *ideal observer analysis* and *energy detector theory*. Signal detection theory estimates the abilities of a system (living organism or machine) to discriminate information-bearing patterns (*stimulus* or *signal*) from random patterns (*noise*) (Green and Swets, 1988). Typical SDT research questions are: What are the fundamental mechanisms that a system uses in signal detection? How the system can adapt to the varying levels of noise that distracts the identification of information signal? Moreover, signal detection theory can be applied to the analysis of the experimental data (e.g. when decision are made under uncertainty). For example, by taking into account the physical and psychological aspects in the decision making process (Green and Swets, 1988).

Ideal observer analysis is a theoretical model of a perceptual system where the "ideal observer" performs a specific information processing task in an optimal way. It can be used to construct a hypothetical model of the perceptual system (e.g. neural network model), and can be used as a reference, when the psychophysical data is analyzed. The ideal observer analysis can be applied also modularly, so that processing at different neural stages is modeled with sub-systems. In the energy detector theory, the sensory system of a human observer can be thought of as a detector that responds when the amount of stimulus energy exceeds the detection threshold (Green and Swets, 1988).

### 2.4.1 Psychoacoustics

Psychoacoustics is a branch of psychophysics that study the perceptual processes of the human auditory system by observing the relation between *sound events* (physical inputs) and *auditory responses* (psychical outputs) (Karjalainen, 2009). However, with appropriate stimulus setup both responses: behavioral (e.g. discrimination threshold) and physiological (e.g. event related fields or potentials) can be collected simultaneously, so that the functionality of the auditory system can be studied more comprehensively. The aforementioned experimental setup can give valuable information of the relation between firing patterns of auditory neurons and the perceptual functions. The potential statistical relationship between the physiological and behavioral data sets can be investigated further with correlation analysis.

The behavioral part of the psychoacoustics rely mainly on the performance measurement experiments, which are used to quantify the sensitivity of hearing (e.g. the threshold of hearing and the dynamic range: Figure 2.1, p. 4). The psychoacoustical measures can be also qualitative that require introspection, such as loudness estimation, sound source estimation, pitch matching or timbre estimation (Karjalainen, 2009).

The *absolute threshold of hearing* (ATH) defines the lower limit of audibility (0 dB) and the threshold of pain the upper limit. The absolute threshold is typically determined with the staircase procedure where the estimate of the threshold is iterated step-by-step by adjusting the sound pressure level. The threshold determination starts with audible SPL, which is then adjusted quieter after each hit (target detected), until the listener misses the target. At that point the staircase reverses so that the stimulus SPL is increased until the subject can hear the stimulus again (hit), which triggers another reversal. The trial is stopped when the reversals are occurring frequently (e.g. 4 consecutive reversals), which means that estimate approaches the absolute threshold asymptotically (Green and Swets, 1988).

The most widely used input signal in psychoacoustical experiments is the *pure tone*, whose sinusoidal periodicity ($p(t) = A\sin(\omega_0 t)$) and mathematical clarity makes it practical to adjust and use as a reference. The duration of pure tone can be changed from transient to continuous-time, and the complexity can be increased by modulating its amplitude or frequency. Moreover, it can be used as a target signal in the tone-in-noise studies (Karjalainen, 2009).

Psychoacoustical studies have proven to be practical in the quantification of the non-linear properties of the human auditory system. The perceived sound intensity (*loudness*) is one of those that varies non-linearly across the time and frequency domain. The *equal-loudness contours* (Figure 2.24, p. 34) illustrate sound pressure levels (dB) that are perceived equally loud over the frequency spectrum. The aforementioned contours are defined usually by loudness matching experiments where subjects are adjusting the intensity with respect to the reference level. The loudness level unit *phon* is fixed according to 1 kHz reference frequency. At 1 kHz frequency, the sound pressure levels and loudness levels are equivalent (Moore, 1995; Karjalainen, 2009).

The *pitch* is a psychoacoustical attribute that describes how we perceive the vibration periodicity of the air with the musical scale. For a sinusoidal pure tone the pitch corresponds approximately to the frequency, and for a periodic complex tone approximately to the fundamental frequency. The dependency of frequency and pitch is logarithmic, so that doubling of the frequency increases the pitch by an octave. The pitch can be perceived if the duration of the tone is long enough ($\geq$ 20 ms). With shorter tones only 'click'-sounds can be perceived (Karjalainen, 2009).

Our ability to discriminate small changes in frequency over time is termed as *frequency discrimination*. The frequency discrimination can be studied by presenting two separate sounds, such as two pure tones sequentially or simultaneously. The just noticeable differences in frequency of a pure tone or fundamental frequency of a complex sound is called as the frequency difference threshold (a.k.a. frequency difference limen). The frequency

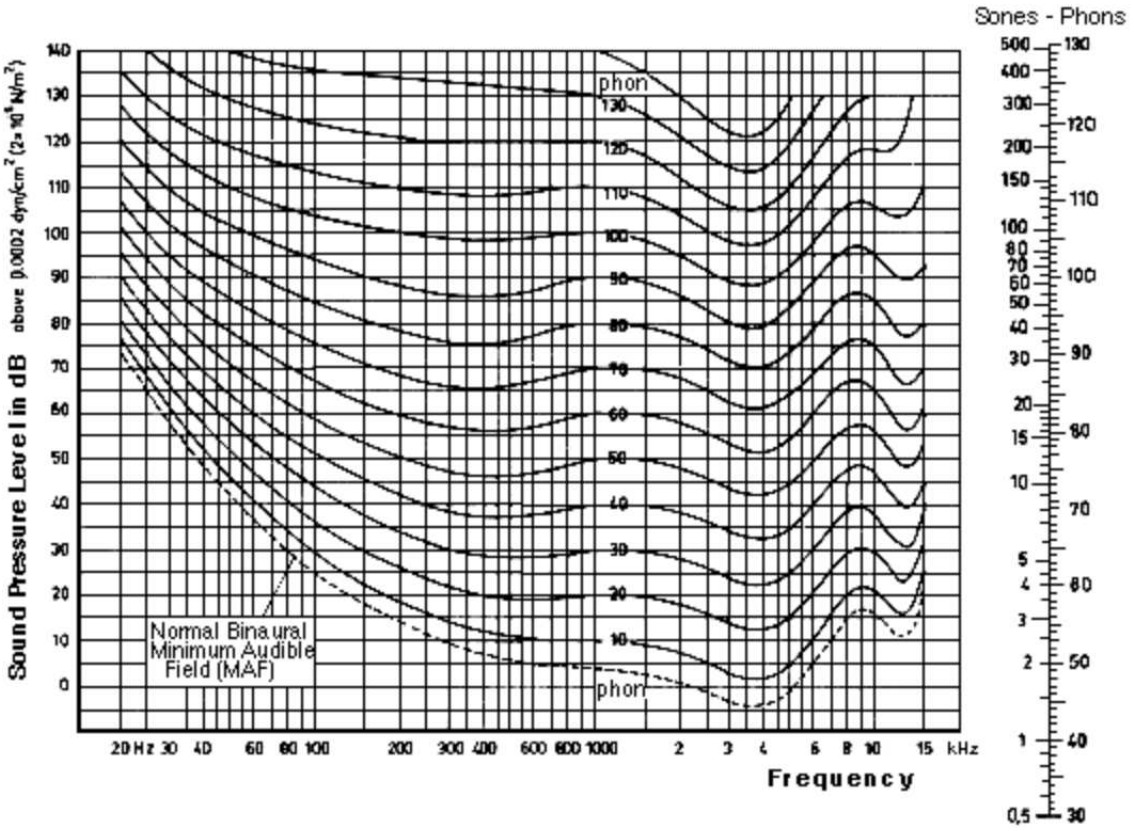



Figure 2.24: Equal loudness contours for pure tones (adapted from ISO recommendation R226).

difference thresholds can be approximated by applying the Weber Equation (p. 32) in frequency domain. The just-noticeable difference is approximately 0.3–1% of the center frequency (Figure 2.25, p. 35), but the frequency discrimination sensitivity varies to some extent, being less sharp at low and high frequencies (Moore, 1995; Karjalainen, 2009). When the differentiated tones are near 1 kHz the just-noticeable difference is between 2–3 Hz.

### 2.4.2 Frequency selectivity and auditory masking

The human auditory system processes complex sounds similarly than the spectrum analyzer. Wideband sounds are decomposed to sinusoidal frequency components in inner ears cochlea, and this spectral information is then encoded to the tonotopic neural representation at the auditory nervous system. Our ability to hear distinct harmonics from the complex sound is based on this tonotopy, and is known as *frequency selectivity* or *frequency resolution*. Thus, if two simultaneously presented tones differ enough in frequency, two separate tones can be heard rather than a combination tone (Ashmore, 2008; Karjalainen, 2009; Rossing et al., 2002).

The resolution of the frequency selectivity depends largely on the processing that happens at the early stage of the human auditory system. Especially, the mechanical properties of

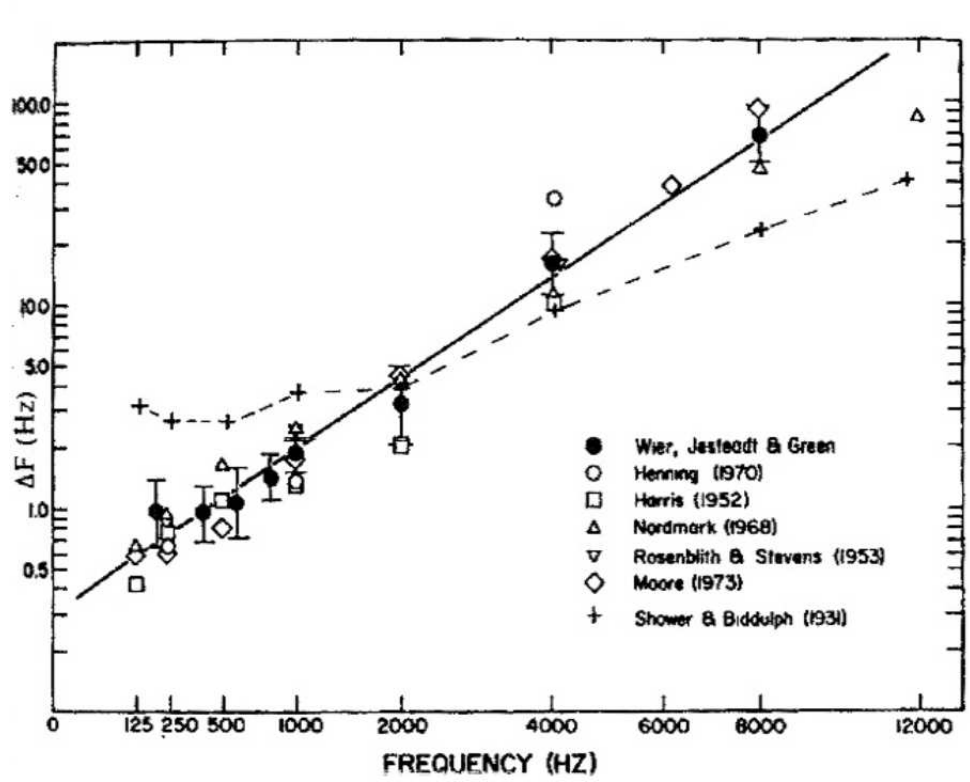



Figure 2.25: JND for pitch (adapted from Moore, 1995).

the cochlea (see Section 2.1.2, p. 8) set limits to the accuracy of frequency information. The complex wideband sounds set the basilar membrane in motion, and the displacement maximums (sinusoidal components) of this unique vibration pattern are decomposed and transferred separately to the higher stages of the auditory system. Similarly than in the Fourier transform, the signal is transformed from the time domain to the frequency domain (Ashmore, 2008).

The filtering or decomposition that happens in the inner ear can be quantified by psychoacoustical masking experiments where the stimulus detection threshold is measured during varying interference signals. The *auditory masking* is a phenomenon where the threshold of audibility of a stimulus (*target*) is raised by another distracting stimulus *masker* (Moore, 1995; Karjalainen, 2009).

The psychoacoustical masking experiments are based on the signal detection and energy detector theories. As was mentioned previously, the signal detection theory estimates the abilities of a human observer to discriminate stimulus from noise. According to the latter theory, the auditory system of a human observer can be thought of as a detector that responds when the amount of sound energy exceeds the detection threshold (Green and Swets, 1988). Since the background noise raises the threshold level of the signal, the phenomenon can be expressed with *signal-to-noise ratio* (SNR) that compares the power of a signal to the power of noise. Because of the wide dynamic range, the signal-to-noise ratio is usually represented in decibel scale (Rossing et al., 2002):

$$SNR_{dB} = 10 \log_{10} \frac{P_{signal}}{P_{noise}} \tag{2.11}$$

The signal-to-noise ratio is not the only attribute that nominates the level of detection threshold. The strength of masking effect depends greatly on, how the power of the masker is distributed in the frequency domain. The masking effect is strongest when the frequencies of signal and masker are close together or overlapping. Based mainly on the functionality of the basilar membrane, there is a certain frequency bandwidth called *critical band* (CB) around the signal within which the perceptual interference is strongest (Moore, 1995; Karjalainen, 2009).

Based on the aforementioned, the frequency selectivity of the human auditory system can be simulated with an array of band-pass filters called *auditory filters*. A single auditory filter passes frequencies within a limited passband (critical band) and rejects frequencies outside that passband. Thus, only noise that contributes in the same critical band with the signal, can increase the masking effect. The filter bank processes the frequency components of input signals separately with the best matching auditory filter(s); with a filter whose center frequency is closest to the frequency of interest [22]. The aforementioned framework is also known as the *power-spectrum model* of masking, since the signal processing is made in the frequency domain, where the short-term temporal fluctuations are omitted (Moore, 1995).

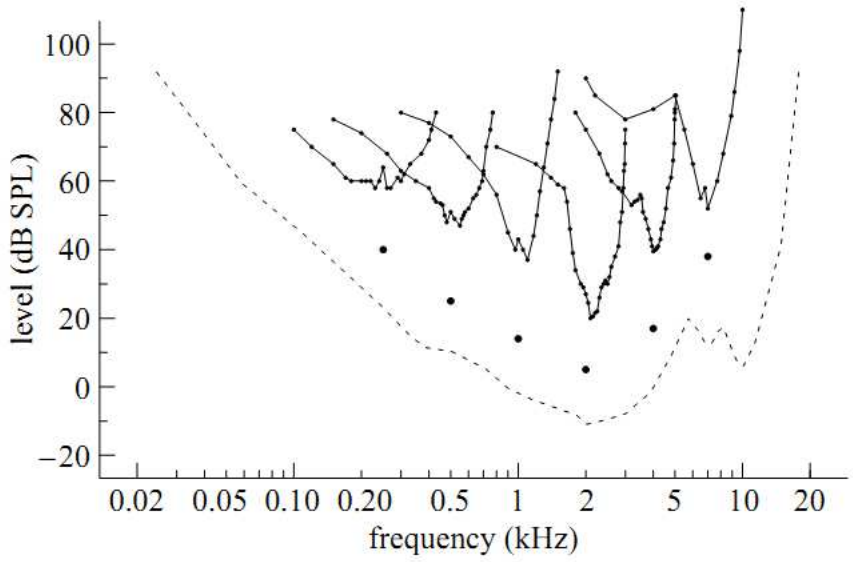


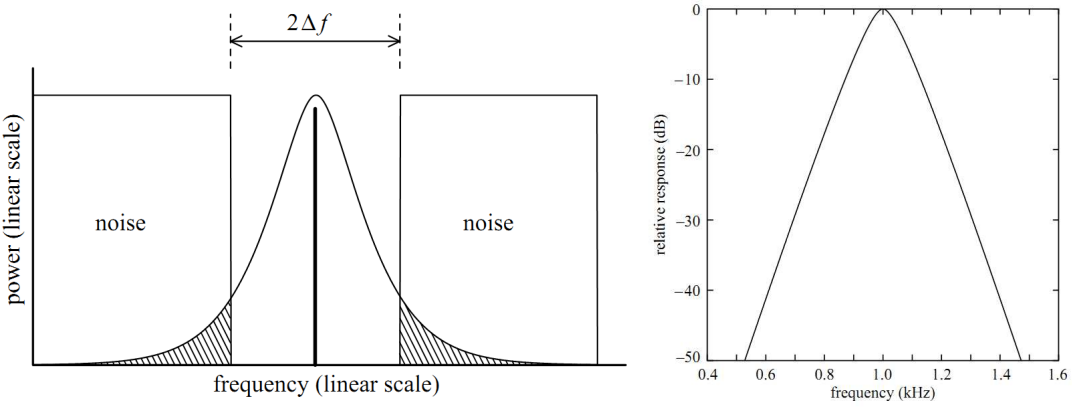

Figure 2.26: PTCs determined with simultaneous sinusoidal masking (black line). The solid circles below the curves indicate the frequency and dB-level of the masker and the dashed line the absolute threshold curve (adapted from Moore, 2008).

[22]The center frequency of an auditory filter correspond to the vibration pattern maximum in the basilar membrane (Ashmore, 2008).

The shape of the auditory filter can be estimated with *psychophysical tuning curves* (PTCs) that are quantified with auditory masking experiments. In such experiment a psychophysical tuning curve is determined at a given centre frequency by setting the signal at low level (*e.g.* +10dB above ATH), and then adjusting the level of the masker so that signal is just masked. The adjustment is made separately with different narrowband maskers, so that frequency range is swept over the test signal. Thus, the SPL of masker at the detection threshod illustrates the shape of the auditory filter as a function of frequency (Figure 2.26, p. 36). By inverting the psychophysical tuning curves we can get the filter bank of auditory filters with overlapping passbands (Moore, 2008). Interestingly, the psychophysical tuning curves resemble the neural tuning curves that are measured straight from the auditory nerves (Figure 2.7, p. 9).

### 2.4.3 Notched-noise method

The shape of the auditory filter can be quantified also with an another masking experiment, which is known as *notched-noise method*. In the notched-noise method the threshold of the signal is determined with respect to the bandwidth of "noiseless spectral notch" that is centered around the signal frequency (Figure 2.27a, p. 37). When the notch width $\Delta f$ around the signal is narrowed, more noise passes through the auditory filter and rises the threshold of the signal. With narrow notches more masking frequencies are interacting within the same critical band with the target signal, which complicates the detection (Moore, 2008). The notch-noise maskers are made by band-stop filtering white noise at the signal frequency. White noise is commonly used in psychoacoustical experiments, because it is a random signal and has flat power spectral density over the range of audible frequencies (Moore, 1995; Karjalainen, 2009).




(a) The probe signal (in the middle) and the masker (on lateral sides)

(b) A typical auditory filter shape

Figure 2.27: Schematic illustration of the notched-noise method (adapted from Moore, 2008).

An example of an auditory filter shape that has been determined with notched-noise method at 1 kHz center frequency is presented in Figure 2.27b (p. 37). The sharpness

of the filter is typically expressed by means of bandwidth of the filter when the response has fallen 3 dB; typically 10 – 15% of the centre frequency (Moore, 2008). The width of the filter affects to the frequency resolution. When the filter is sharp, only the frequencies near the center frequency pass through the filter. The height of the filter affects to the level of the output – high filters gain the inputs near the center frequency.

However, the effect of masking is not as symmetrical as the notched-noise method predicts. For example, low frequency sounds put the entire basilar membrane in vibration, and thus, the masking lowers also the detection of high frequencies. Moreover, the intensity has effect on the masking symmetry. At low and moderate noise levels the auditory filters seem to more symmetric than at high noise levels (Moore, 1995). Sometimes it can be more practical to simulate the masking with a simplified model, that takes only into account the bandwidth of the filter. The *equivalent rectangular bandwidth* (ERB) model represents the auditory filters with rectangular band-pass filters. At low and moderate sound levels ERBs can be approximated with the equation:

$$ERB = 24.7(0.00437f + 1) \tag{2.12}$$

, where the $f$ is the center frequency in Hz (Moore, 2008). With 1 kHz center frequency the Equation 2.4.3 (p. 38) gives 132 Hz as equivalent rectangular bandwidth.

Instead, the critical band around the center frequency is typically approximated with the following equation (Zwicker and Terhardt, 1980):

$$CB = 25 + 75 \times (1 + 1.4f^2)^{0.69} \tag{2.13}$$

At 1 kHz the Equation 2.4.3 (p. 38) gives 160 Hz as critical bandwidth. The critical bandwidth for frequencies above 1 kHz is typically about 10-17% of the center frequency (Moore, 2008).

The notched-noise method can be applied to the neuroscientific research by measuring the tone-evoked magnetic fields (MEG) or potentials (EEG) non-invasively over the cerebral cortices during the notched-noise expriment. The cortical responses can be evoked by a sequence of pure tones that are payed together with the continuous-time notched-noise maskers. When the masker notches are parametrically narrowed around the signal in the notched noise experiment, the N100(m) components are suppressed in amplitude. The suppression of N100(m) components is strongest in conditions where the spectrum of the masker is within the critical bandwidth of the signal frequency (Hari and Mäkelä, 1988). The neural data can be used as a function where auditory evoked response (e.g. N100m or SF components) amplitudes are mapped to the notch width of the noise maskers:

$$f(\Delta F) = A_{AEF} \tag{2.14}$$

where $\Delta F$ is the notch width of the noise masker [Hz] and $A_{AER}$ is the amplitude (e.g. [fT/cm]).

## 2.5 Attention

Attention is typically defined as a cognitive process that controls and optimizes the perceptual processing by allocating the limited neural processing capacity to the most critical ones (Lachter et al., 2004). It has been compared to a spotlight that illuminates the stage, so that drama is displayed brightly – however, there may be insignificant characters in the shadows. A more technical analogy for attention is a dynamic filter system that selects task-relevant sensory inputs from the environment and filters out the non-relevant and distracting inputs.

Attention modifies the sensory perception both in automatic and contorolled manner. Automatic process is triggered from the bottom when salient or surprising sensory events catch our attention. Because of this automatic and proactive nature, this *bottom-up* process is also known as *involuntary attention* or *pre-attention*. *Voluntary attention* is controlled and goal-orianted activity, where the process start from the top. Cognitive *top-down* factors, such as knowledge, expectation and current goals control the neural processing and representation at lower stages. Voluntary attention can organize multiple brain areas to operate consistently and efficiently in order to carry out the goal-oriented actions. Voluntary attention is also called as *selective attention*, which emphasize conditions where the environment is full of sensory distractors and the attentional task requires selection (Knudsen, 2007).

Attention works in all sensory domains and can be directed voluntarily to a single modality, which is known as *unimodal attention*. Attention improves the sensory performance of the attended modality by shortening reaction times and enhancing stimulus discrimination (Murray, 2005). Attention can operate also coherently between modalities (*cross-modal attention*), if the cognitive task requires multimodal sensory processing (Driver and Spence, 1998), such as linguistic functions do.

The functionality of attention can not be explained without a storage system that preserves impressions after the physical stimuli have decayed. *Sensory memory* is a concept of transient storage that keeps sensory events automatically in a buffer from milliseconds to couple of seconds, so that this information is available for further processing (Lachter et al., 2004). It is presumed that each sensory modality has a corresponding memory system, such as *echoic memory* for auditory events and *iconic memory* for visual events (Darwin et al., 1972).

*Short-term memory* (STM) and *working memory* (WM) are concepts of dynamic memory, which are often used interchangeably since their basic functionality is substantially the same (Aben et al., 2012). The basic functionality of these systems (STM and WM) is to store information (4 – 7 items) for a brief period of time (i.e. 18s – 30s), so that the information is available for detailed analysis and manipulation. Cowan (1998) suggest that WM is only an extended version of STM that is connected to *long-term memory* (LTM). Long-term memory is a concept of a quasi-permanent memory, where information can be stored indefinitely, and retrieved when it is needed.

The conceptual framework of attention (Knudsen, 2007) presents the main components

that have also contribution on the function of sensory system, especially on the behavioral sensitivity and performance. The main components: working memory, competitive selection, top-down sensitivity control, and bottom-up salience filters (Figure 2.28, p. 41) operate in a recurrent loop and optimize the sensory output (neural representation). Salience filters automatically enhance the relevant inputs (e.g. infrequent, instinctive or learned stimuli) from the environment. Competitive selection pick up the most relevant stimuli and give them access to the working memory. Working memory is a temporary storage for information that is used in goal-oriented actions. Sensitivity control is a top-down process that adjusts the relative strengths of the different information streams, which in turn compete for access to the working memory. The recurrent loop adjusts the neural representation, which is the basis of the improved behavioral sensitivity and performance in sensory perspective (for a more inclusive review, see Knudsen (2007)).

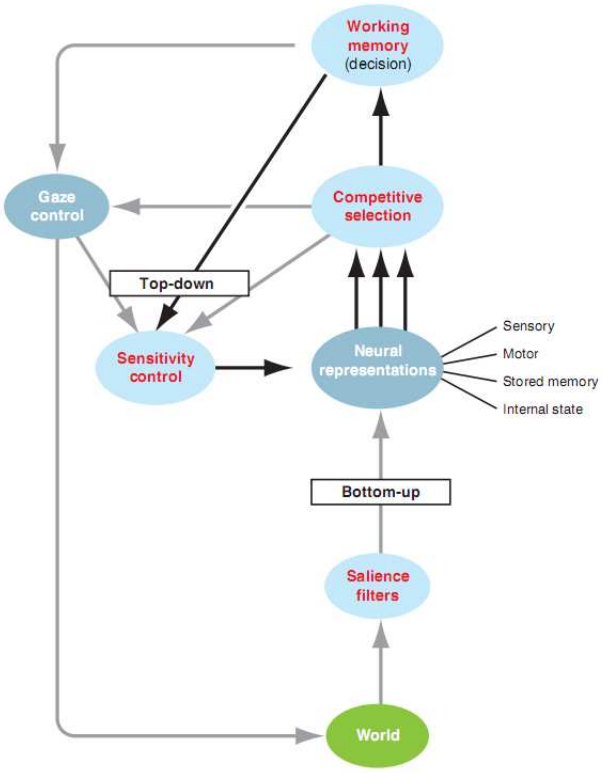



Figure 2.28: A conceptual framework of attention (adapted from Knudsen, 2007).

The experimental research of attention is based heavily on the functional imaging studies that commonly try to map, how neuroanatomical structures are linked to the attentional functions. The best spatial resolution can be obtained by fMRI and PET techniques, but event-relelated response studies require temporal resolution that can be achieved only by MEG or EEG. One way is to combine the data that is acquired by MEG/EEG and fMRI/PET measurements (Luck, 2005).

The transient effects of attention can be studied with MEG or EEG by using an *oddball procedure* where infrequent *deviant* stimuli are used in a series of *standard* stimuli. The deviants, such as a sudden change in stimulus intensity or frequency are used to provoke

divergent responses. A typical deviant evoked response is a negative difference wave that peaks between 150–250 ms after the stimulus onset. The amplitude (missmatch negativity) and latency is suggested to reflect differences of neural representations between deviant and standard stimulus (Näätänen et al., 1978), or the negative component is suggested to reflect the bottom-up mechanisms that gate novel sensory events to the awareness (Jääskeläinen et al., 2004). The effect is suggested to relate the *stimulus-specific adaptation* (SSA) – a suppression effect on standard-evoked N100(m) components that is caused by stimulus repetition (May and Tiitinen, 2010).

The oddball paradigm can be also used to study top-down mechanisms of attention, so that the deviants are used to keep focus on auditory task, but only the standard-evoked responses are used in the data analysis. A typical modulatory effect of top-down attention is a *negative difference* (Nd) pattern, which is also termed as attention-related *processing negativity* (Hansen and Hillyard, 1980; Näätänen, 1990; Woods et al., 1994). The early phase of auditory evoked Nd wave typically shows up between 100–300 ms from stimulus onset and could overlap with the N100 or with the N100-P200 complex. The late phase of Nd wave occurs roughly between 300–400 ms and is located more frontally at the cortex (Hansen and Hillyard, 1980).

### 2.5.1 Neuroanatomy of attention

In neuroanatomical viewpoint, the human attentional system is not a single entity; instead, multiple neural sites co-operate and implement different but interrelated functions. These operational networks form unified systems, whose neuroanatomy differ according to the attention-related functionality (Posner and Petersen, 1990). Posner and Petersen (1990) separated three attentional networks for *alerting*, *orienting* and *detecting*. However, the three-piece model was revised afterwards by Posner and Rothbart (2007), and the detecting subsystem was replaced by *executive control*.

The anatomy of alerting network includes frontal and parietal cortical areas and subcortical areas in thalamus (Figure 2.29, p. 43). The orienting network has subcortical structures; pulvinar nucleus in thalamus and superior colliculus in the midbrain, as well as cortical structures; *superior parietal lobe*, *temporal parietal junction* and *frontal eye field* (FEF). The anatomy of executive network comprises cortical sites; the basal ganglia, *anterior cingulate* and areas in frontal cortex. The function of the aforementioned networks is modulated by neurotransmitters, such as norepinephrine (alerting), acetylcholine (orienting) and dopamine (executive) (Posner and Rothbart, 2007).

The neuroanatomy of sensory attention can be divided hierarchically to neural sites that control attention and the target sites that are modulated by attention (Yantis, 2008). The prefrontal cortex (PFC) has a central role in the top-down control of attention, especially the intraparietal cortex and superior frontal cortex (Corbetta and Shulman, 2002). The prefontal cortex has myriad projections to the sensory cortices, by which it can enhance the processing of relevant stimuli and inhibit the processing of distracting stimuli (Arnsten et al., 2009). Moreover, neural structures that are associated to attentional control have

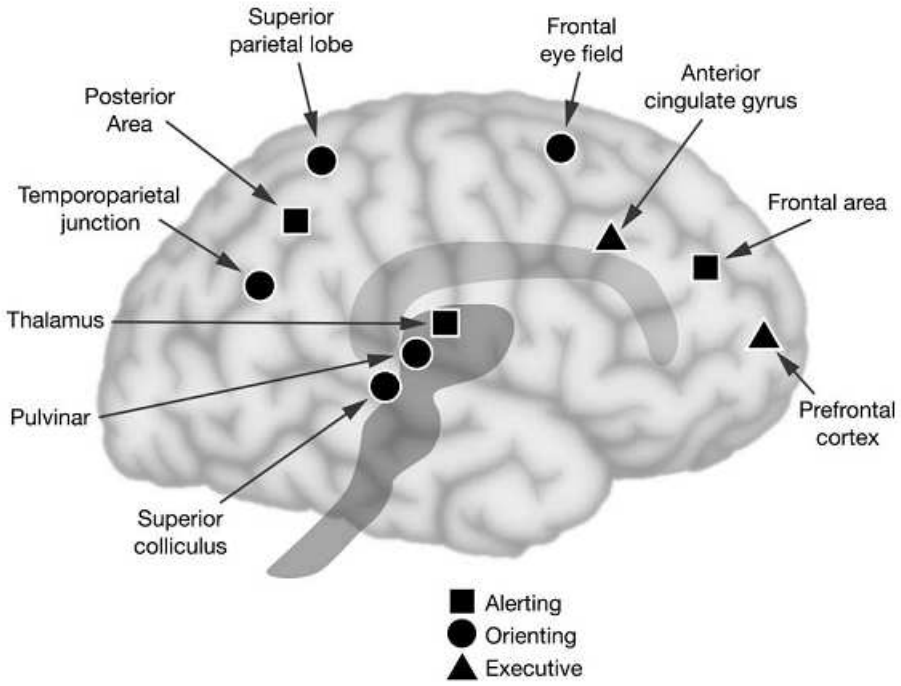



Figure 2.29: The anatomy of the three attentional networks: alerting (squares), orienting (circles) and executive control (triangles) (adapted from Posner and Rothbart, 2007).

been localized in frontal eye field, superior parietal lobule and intraparietal sulcus (Yantis, 2008).

The study of Corbetta and Shulman (2002) found evidence of two anatomically and functionally distinct neural systems that regulate attention (Figure 2.30, p. 44). The *dorsal fronto-parietal network* was involved in voluntary top-down controlled attention and was distributed anatomically to *intra-parietal sulcus/superior parietal lobule* (IPS/SPL) and *frontal eye field* (FEF). This network was active during visual tasks, when subjects attended selectively on stimulus feature or location. The *ventral fronto-parietal network* was involved in bottom-up triggered attention, for example, when salient or unexpected events catch attention. The anatomical locations of ventral fronto-parietal network were in *temporo-parietal junction* (TPJ) and areas of the *inferior frontal gyrus/middle frontal gyrus* (IFG/MFG) (Corbetta and Shulman, 2002).

Interestingly, similar dorsal and ventral networks were active during top-down controlled and bottom-up triggered auditory attention (Salmi et al., 2009). This is logical since the attentional functions are typically cross-modal (e.g. attention related linguistic functions). The study of Salmi et al. (2009) showed that top-down controlled auditory attention enhanced neural activity IPS/SPL and FEF, but also TPJ and IFG/MFG, which are typically associated to bottom-up triggered attention.

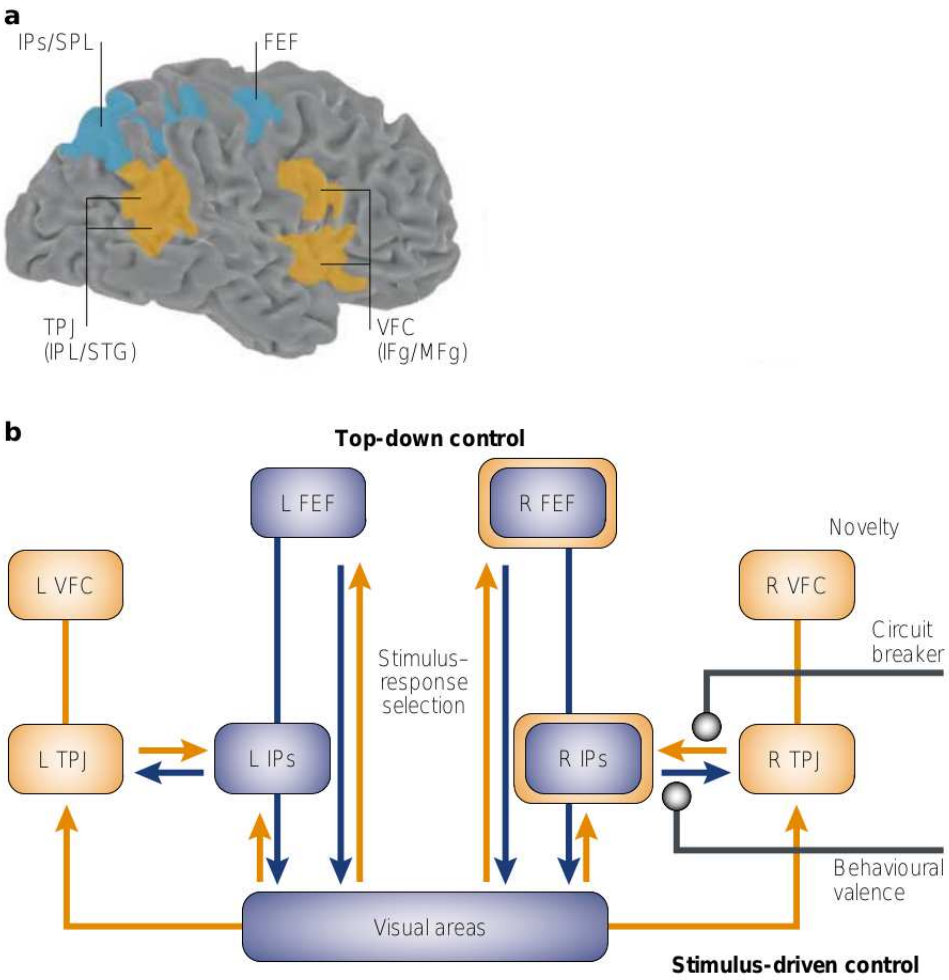



Figure 2.30: Neuroanatomical model of attentional control. Dorsal fronto-parietal network (blue) and ventral fronto-parietal network (orange) (adapted from Corbetta and Shulman, 2002).

### 2.5.2 Neurophysiological mechanisms of attention

The processing at sensory system is based on hierarchy, where neurons with similar sensitivity properties are mingled together into functional ensembles (see Section 2.2.3, p. 15). Selective attention can improve the behavioral sensitivity and performance by affecting on those neural ensembles that are processing the attended object or feature. Thus, selective attention can modulate neural processing at multiple stages in central nervous system (Kandel et al., 2000). For example, the modulatory effects of selective auditory attention have been detected in numerous locations in auditory system: in the auditory pathway, primary and secondary auditory cortices, and even at the level of cochlea (Giard et al., 2000). The modulatory effects of selective auditory attention has been located especially in the supratemporal plane of the auditory cortex (Rif et al., 1991; Woldorff et al., 1993). Moreover, neural sites outside the auditory system have showed also increased activity under auditory attention: frontal cortex, bilateral pre-central and left post-central cortices, the supplementary motor area, and the right mid-thalamus (Giard et al., 2000).

The *sensory enhancement theory* assumes that the neural mechanism of selective attention is based on enhancements in the neural level representations of the attended objects. This would strengthen the contrast between relevant and non-relevant objects and ease the selection. The sensory enhancement theory has been investigated largely with event-related response experiments, where the responses were compared between attended and unattended conditions. The fundamental finding of these studies was that selective attention gained the response amplitudes, which further led to term *gain model* to describe the neural mechanism. The gain model assumes that selective attention strengthen responses to attended stimuli and suppress responses to non-attended stimuli, which leads to amplification of the cortical responses (Desimone and Duncan, 1995).

The gain based effects have been observed in both auditory (Hillyard et al., 1973; Rif et al., 1991; Woldorff et al., 1993) and visual (Desimone and Duncan, 1995) modalities. Single-cell recordings demonstrated that object-based attention increased the baseline activity (i.e. firing rates) in the auditory cortices (Hocherman et al., 1976) and visual cortices (Roelfsema et al., 1998) of the macaque monkey. In humans, fMRI study (Grady et al., 1997) showed attention-related enhancement in primary and secondary auditory cortex.

In auditory modality the gain based amplifications have been detected at early latencies: P20(m) and P50(m) components (Woldorff and Hillyard, 1991; Woldorff et al., 1993), and at late latencies: N100(m) (Hillyard et al., 1973; Rif et al., 1991; Woldorff et al., 1993), P200m (Rif et al., 1991) and P300 (Hillyard et al., 1973) components. The gain based amplitude increments were found also on the sustained response components: EEG study (Picton et al., 1978) and combined EEG/MEG study (Arthur et al., 1991).

However, the gain model can not explain, how the human auditory system can segregate the task-relevant signals from noise, if the masking frequencies are overlapping with the critical band of signal. The aforementioned condition corresponds to a notched-noise experiment, where the notches are narrowed within the critical band of signal (see Section 2.4.3, p. 37). In such overlap condition, the neural gain would amplify both, the signal and the masker since both inputs are processed by same the receptive fields at tonotopic cortices. However, frequency-specific selection succeeds also in conditions where the masking frequencies are expanded within the critical band of signal, which indicates that the frequency resolution can sharpen during attention. The improved resolution (a.k.a *selectivity increase*) has been explained by a concept *short-term plasticity*; top-down (or bottom-up) controlled neurophysiological changes that show up in a time scale of milliseconds to minutes (for a review, see Jääskeläinen et al. (2007)).

The *neural tuning* model complements the "gain-only-model" by receptive field placticity that can sharpen neurons feature selectivity. This means that the auditory system can adapt more diversely to the attentional demands, by improving the signal-to-noise ratio of the targets. The tunign model explains better our ability to resolve meaningful information from noise. Functional MRI and electromagnetic (EEG/MEG) measurements (Ahveninen et al., 2006; Altmann et al., 2008; Ahveninen et al., 2011) have found evidence of the neural tuning model. These studies suggest that besides the gain based effects, selective auditory attention can also sharpen the feature selectivity. At cortical level the sharpening can be based on lateral inhibition (i.e. center-excitation and surround inhibition), which

could reduce the overlap between neural populations that represent different frequencies (Jääskeläinen et al., 2007).

### 2.5.3 Frequency-specific selective attention

The Broadbent's selective filter theory (Broadbent, 1958), represents the function of selective attention with dynamic filters that constrain the sensory stream according to the focus of attention. The idea of attentional filters can be combined to the auditory filter concept (see section 2.4.2, p. 36), where the frequency resolution of human auditory system is simulated with a filter bank of band-pass filters. A single band-pass filter allows a range of frequencies within the critical bandwidth to pass through, while frequencies outside the cut-off frequencies are stopped. In the *dynamic auditory filter synthesis* the frequency-specific selective attention adjusts the auditory filters dynamically (Figure 2.31, p. 46) by stretching the height of the filter (gain) and by narrowing the width (CB) of the filter (selectivity increase). The aforementioned dynamic auditory filters framework actually illustrate the function of the neural tuning model, and sums up the experimental paradigm of the current research.

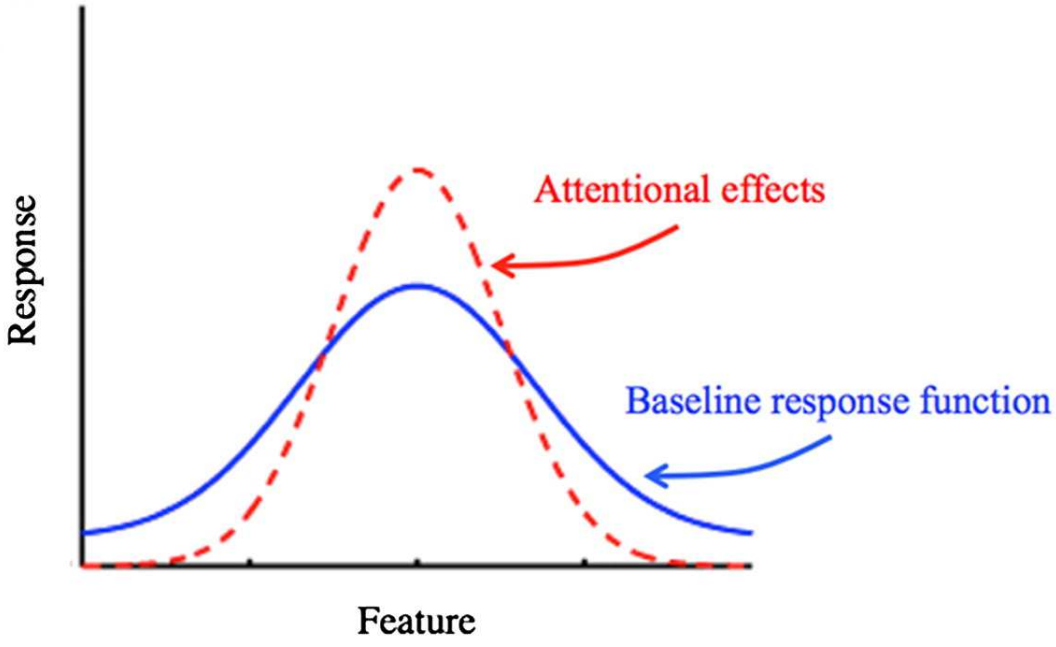



Figure 2.31: Attention as an increase in gain and sharpening of tuning curve (adapted from Hickok et al., 2011).

The hypothetical filters can be investigated with physiological and behavioral measures; by a notched-noise experiment where the threshold of a fixed sinusoidal signal is measured as a function of the bandwidth of a passband noise masker (see Section 2.4.3, p. 37). The sinusoidal signal can be the target in behavioral experiment, but also the stimulus that provoke the auditory evoked fields in the MEG measurement.

When the masker notches are parametrically narrowed around the signal in the notched noise experiment, the detection of probe tones becomes more difficult (Moore, 2008). This reflects also to the N100(m) components, which are suppressed in amplitude when the masker notch is decreased. The suppression is strongest in conditions where the spectrum of the masker is within the critical bandwidth of the signal frequency (Hari and Mäkelä, 1988). With a suitable psychophysical experimantal framework, the neural (tone-evoked responses) and behavioral responses (discrimination threshold) can be measured simultaneously during varying noise conditions, and thus, the responses are comparable. Also the effect of auditory attention can be investigated with the same experimantal framework, by measuring responses with identical stimulus setup during auditory attention and during control condition.

## 2.6 Aims of the research

The basic idea in the current study was to apply two conceptual frameworks: auditory filters (Fletcher, 1940; Patterson, 1986; Glasberg and Moore, 1990) and selective filters (Broadbent, 1958) to the research of selective auditory attention. The former concept simulates the frequency resolution of the human auditory system, and the latter, the dynamic property of selective attention to constrain the auditory stream according to the focus of attention. These are simplified models, but similar filtration is suggested to happen also at the auditory cortices where neural ensembles process frequency-based information under selective auditory attention.

The hypothetical model of the current study (Synthesis 2.5.3, p. 46) simulates the functionality of the frequency-specific selective attention with auditory filters (Figure 2.31, p. 46) that are adjusted dynamically by stretching the height of the filter (gain) and by narrowing the width (CB) of the filter (selectivity increase). The aforementioned framework illustrates the neural tuning mechanism that is suggested to be the key mechanism, when auditory attention is used selectively to segregate frequency-based information under difficult noise conditions (Kauramäki et al., 2007; Okamoto et al., 2007b). The recent studies suggest that this neural tuning is based on short-term plasticity changes in the secondary auditory cortices, where receptive fields are tuned (re-shaped) to the attended features (Ahveninen et al., 2006, 2011; Altmann et al., 2008).

The primary aim of the research was to investigate the neural mechanisms of selective auditory attention with physiological and behavioral measures, especially in conditions, where the notched-noise maskers were narrowed within the critical bandwidth of the target tones ($\leq$ 160 Hz). The secondary aim was to show evidence of the neural tuning and the improved frequency selectivity in the narrow-notch conditions.

The specific research hypotheses were:

(a) Narrowing of the notch width of the notched-noise masker impairs the pitch discrimination sensitivity (d') and suppresses the amplitudes of the auditory evoked N100m and SF components.

(b) The masker-induced suppression gets stronger when the masker notch is narrowed within the critical band of target tones ($\leq$ 160 Hz).

(c) The modulatory effect of selective attention is more powerful in narrow notch conditions, when the masker notches are within the critical band ($\leq$ 160 Hz).

# Chapter 3

# Materials and methods

The experimental paradigm was based on the following assumptions: 1) The ability to discriminate small changes in sound frequency is related to the quality of the human auditory system, which is organized tonotopically. 2) The tonotopic structures have been located in the primary auditory cortex and the belt areas (Talavage et al., 2004; Humphries et al., 2010). 3) The frequency resolution of the human auditory system can be modeled with auditory filters, and these filters can be investigated with psychophysical notched-noise experiments[1] (Fletcher, 1940; Patterson, 1986; Glasberg and Moore, 1990). 4) The tone-evoked magnetic fields can be measured non-invasively over the cerebral cortices, since population-level activity of pyramidal cells generate extracranial signals (Hämäläinen et al., 1993). 5) Selective attention to tones strengthen the neural activity in auditory cortices mainly at secondary auditory areas (Woldorff et al., 1993; Petkov et al., 2004). 6) The prevailing theories of the neurophysiological mechanisms suggest that attention both gain (Hillyard et al., 1973; Rif et al., 1991; Woldorff et al., 1993) and increase the selectivity (Ahveninen et al., 2006, 2011; Altmann et al., 2008) of task-relevant inputs at auditory cortex.

## 3.1 Experimental setup

The modulatory effects of selective auditory attention were investigated with experimental setup where auditory evoked magnetic fields were measured with MEG during a tone-in-noise experiment. The measured AEF components – N100m onset fields (latency $\sim$ 100 ms) and sustained fields (SF; latency $\sim$ 200–600 ms) were evoked by pure tones embedded within continuous-time notch-filtered noise maskers. The both components were calculated from a fixed subset of 6 planar gradiometer pairs selected mirror-symmetrically over the both temporal lobes. During the tone-in-noise experiment two parameters: the focus of attention – auditory task vs. visual control task – and the notch width of the

[1]The threshold of a fixed sinusoidal signal is measured as a function of the bandwidth of a passband noise masker.

band-stop filtered spectral notch centered at the signal frequency were varied in order to monitor how these parametric changes reflect to the processing at auditory cortices.

The focus of attention was controlled by instructing subjects to identify barely audible frequency deviants 1020-Hz target tones (P=0.1) from more frequently occurring 1000-Hz standard tones (P=0.9), or respectively focus attention on a visual control task where identical auditory stimuli were playing. Because the auditory stimulus setup was identical in both, auditory and visual task conditions, the effect of auditory attention could be extracted by comparing the responses between conditions. The difficulty level of the frequency discrimination task was varied by seven different notch widths (P2) between $\pm$500 and 0 Hz. The detection sensitivity ($\sim$ hit rates) and reaction times were measured as the behavioral data representing the auditory performance. The neural data were used as a function where N100m and SF amplitudes were mapped to the notch width of the noise maskers (Function 2.14, p. 38) and compared it between task conditions. The behavioral data were used in parallel with MEG data to investigate the relation between auditory cortex activity and auditory performance.

## 3.2 Subjects

Eighteen healthy voluntary university personnel and students participated in the psychophysical experiment. Sixteen of them where right-handed and two were left-handed. The majority of subjects were native Finnish speakers, exept two candidates were speakers of native English and native Russian. Fourteen subjects (N=14) – ten male and four female subjects (71% male and 29% female) – were included in the data analysis. Four subjects were excluded from the analysis due to technical difficulties (e.g. excessive number of artefacts or experiment was aborted due to claustrophobia). The subjects were aged between 21 and 46 years ($\overline{x} = 29.2$, $\sigma = 6.1$ years). All participants had normal hearing and vision, or alternatively the vision was corrected to normal with contact lenses. The permission to MEG experiment was approved by the ethics committee of the Helsinki and Uusimaa Hospital District. The subjects were not paid for the participation to the experiment and the voluntariness was confirmed with a written agreement.

### 3.2.1 Stimuli

The transients in the magnetic fields were evoked by auditory stimuli which were embedded within the continuous-time notched-noise maskers (Figure 3.1, p. 52). In the behavioral task the subjects were instructed to attend either on sound frequency (*auditory attentive condition*; AttAud) or on the visual control task (*visual control condition*; AttVis). In the AttAud condition subjects tried to discriminate pure tones of higher frequency (1020 Hz *deviants*, P=0.1) from more frequently occurring pure tones (1000 Hz *standards*, P=0.9). The auditory stimuli were 300-ms sinusoidal tones with 5-ms onset and offset ramps. The inter stimulus interval between stimuli was randomized at range 1800–2200 ms. The auditory stimuli were adjusted to the detection threshold in order

to extract the attentional impact on AEFs. Too high sound intensities would otherwise cause the responses to saturate to their maximum values and thus overpower the effect of attention.

The behavioral task was divided in eight blocks, where seven of them had notched-noise masking on the background and the eight block was reference without masking. Each notched-noise blocks had different spectral notch around the auditory stimuli. The level of frequency discrimination task was at easiest in blocks where the spectral notch around the target tone was wide ($\leq \pm 500$ Hz) and more difficult when the notch was reduced. In the most difficult condition the notch was reduced to 0 Hz, which equals pure white noise and the frequency discrimination was made intentionally "impossible". The notched-noise maskers were made by band-stop filtering the white noise at 1 kHz center frequency so that the filtered noise contained a "notch" in the frequency band (Figure 3.1). Auditory stimuli were prepared in Matlab (R14, Math-Works Inc., Natick, MA, USA) using 16-bit quantization and a 48 kHz sampling rate. Seven different maskers were prepared by bandstop filtering Gaussian white noise with symmetrical stopbands $\pm$500Hz, $\pm$300 Hz, $\pm$200 Hz, $\pm$150 Hz, $\pm$100 Hz, $\pm$50 Hz and 0 Hz around 1000 Hz. For instance, $\pm$200 Hz corresponds to a 800–1200 Hz stopband. The attenuation in the stopband was 120 dB and the transition bands were very steep. Standard and deviant auditory stimuli were randomized and concatenated to a continuous 8-minute binaural audio-files (WAV) with a 44.1 kHz sampling rate and 16-bit quantization.

In the AttVis condition the subjects tried to discriminate deviants of higher spatial frequency from standard spatial frequency. Visual stimuli (Figure 3.1, p. 52) were wavelet patterns, known as Gabor patches. A Gabor patch is a uniformly oriented grating, whose luminance profile is sinus. The deviant patterns (P=0.1) got a slightly higher spatial frequency than the standard patterns (P=0.9). Both stimuli were presented for 400 ms with a constant 1.8 second interstimulus interval. The standard and deviant patterns were concatenated in a random order to a continuous video-file. The audio-files were played simultaneously with the audio-files in the experiment. The visual task was used as a control instead of passive condition so that the arousal-related effects could be minimized. The difficulty levels were customized to be roughly equal between conditions so that the cognitive load was kept up high during the experiment. The focus of attention was also switched between auditory task and visual control task in turns (2-minute periods) to minimize the bias in the alertness between conditions.

## 3.3 Instrumentation

### 3.3.1 Audiovisual system

Sounds and animation were played with Presentation software (v12.0, Neurobehavioral Systems Inc., Albany, CA), which was running on a computer located outside the shielded room. Auditory stimuli were delivered to the shielded room via an electrostatic panel speaker (Panphonics SSH-SQW sound shower, Panphonics, Espoo, Finland), which was

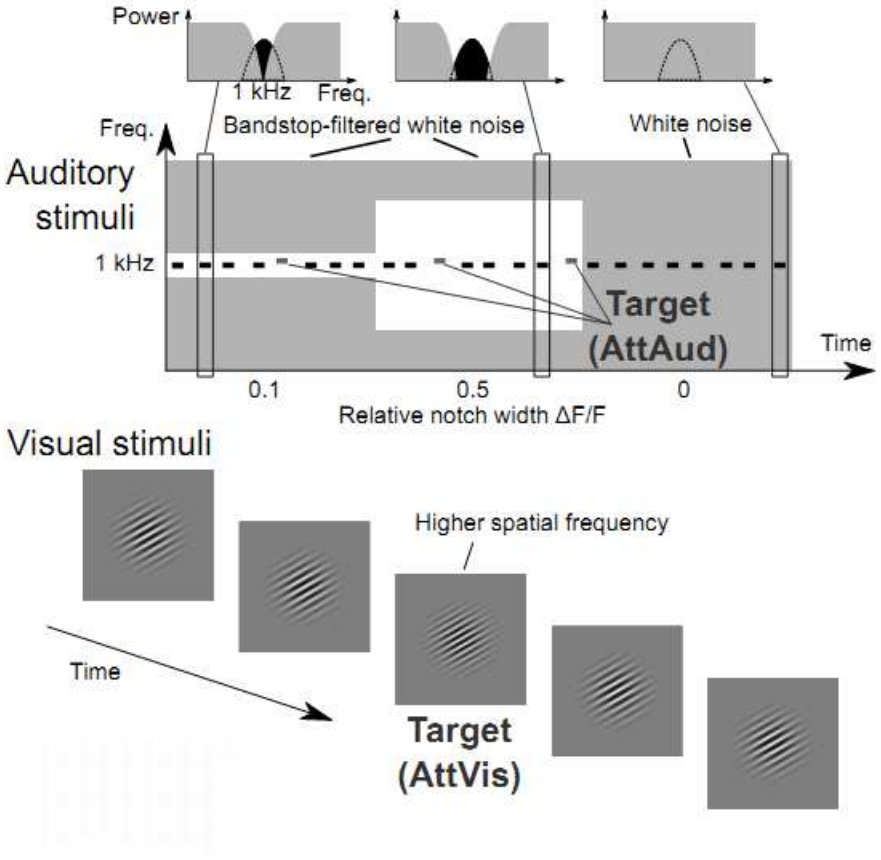



Figure 3.1: Auditory and visual stimuli [adapted from Kauramäki et al., 2012].

mounted on the top corner of the wall and was directed towards the subject's face. The distance from the speaker to the subject's head was about 2.8 m. The dimensions of the panel speaker were 60x60 cm and the frequency band 400 – 16 000 Hz (-6 dB/oct.). The electrostatic loudspeaker elements are suitable for MEG measurements because the drive currents are rather low and no magnets are involved. Thus electrostatic panel speakers do not interfere with MEG signals.

The video signal was transmitted to the shielded room from the outside with a video projector. The picture was projected through an opening in the wall to a 30.4" semitransparent back-projection screen. The screen was placed in front of the subject at a distance of 1.5 m.

### 3.3.2 MEG system

The MEG-measurements were executed at Low Temperature Laboratory, Aalto University School of Science and Technology. The MEG system used in the experiment was a 306-channel whole-head neuromagnetometer, Vectorview Elekta Neuromag system (Elekta Oy, Helsinki, Finland). The 306 channels consist of 102 units where each comprises a

magnetometer (Bz) and two orthogonal planar gradiometers (dBz/dx and dBz/dy). The MEG equipment was placed in a three-layer magnetically shielded room, which was covered with two layers of mu-metal and single layer of aluminum to isolate the measurement environment from the magnetically hostile environment.

All of the 306 MEG channels were used in the data acquisition. A digital trigger channel was used to track the on-set times of auditory stimuli and the corresponding neural responses. Also an additional EOG (electro-oculography) channel was applied to exclude epochs contaminated by eye movements or eye blinks.

### 3.3.3 MEG data acquisition

The ERF data were collected with Neuromag software (Neuromag Ltd, Espoo, Finland) provided by the equipment manufacturer. The MEG data acquisition software was running on HP Unix 64-bit PC. The acquired MEG signals were sampled at a 2 kHz sampling rate and were filtered with 0.1–650 Hz bandpass (online-averages). Rejection criteria was applied to remove probable extracranial artefacts. Epochs were automatically rejected from the average if any of the following magnitudes were exceeded: 3000 fT/cm in the gradiometer channel, 400 fT in the magnetometer channel or 150 $\mu$V in the EOG channel. The criteria was used both in the case of the online-averages and raw fiff-files (offline-averages). The online-average was used during the data acquisition to monitor that enough good segments (N>100) were collected to the data set in order to acquire satisfactory SNR. However, all MEG data including raw fiff-files and online averages were stored on a hard disk for further treatment. Also the behavioral data which contained information of hit rates, reaction times and false alarms were stored. In addition, the data from the digital trigger channel were saved because it contained sound onset information that was used in synchronization of ERFs and the corresponding behavioral responses. The microphone channel information was used only for internal purposes to secure that the trigger-to-stimulus jitter was good enough.

## 3.4 The course of the experiment

### 3.4.1 Subject preparation

Before the actual measurements several preliminary preparations were made to the subjects including: demagnetization of the subject (1–5 minutes), attachment of the head position indicator (HPI) coils and EOG electrodes (5 minutes), digitizing the anatomical landmarks and HPI coil locations (1–5 minutes), seating the subject and giving instructions (1–5 minutes).

The subject demagnetization was made by taking out all objects that contain metal: eyeglasses, earrings, hairpins, necklaces, piercings on the head, wrist watch, underwire bra, belt and shoes. If the subject was not able to see the necessary instructions on the screen

without eyeglasses, the vision was corrected by the use of contact lenses. Clothes that included magnetic buttons, rivets and zippers were replaced with the laboratory's non-magnetic clothes. The magnetization was tested case-by-case if the subject had a possible artefact source, such as make-up, hair dye, dental work or a metal implant.

After the demagnetization EOG electrodes and HPI coils were attached to the subject's head. One of the EOG electrodes was placed below the left eye and the other on the left temple near the canthus. The EOG ground electrode was attached on the center of the forehead. Two of the HPI coils were attached widely apart on the forehead near the hairline and two were attached behind the earlobes as high up as possible. The HPI coils were placed high enough to be covered by the MEG sensor array because during the measurement they are utilized to track the head position with respect to the sensor array (co-registration).

Thereafter the anatomical landmarks and the HPI coil locations were digitized by using an Isotrak 3D-digitizer. The digitization procedure started from anatomical landmarks – the left preauricular point, the nasion and the right preauricular point – and continued to the HPI coil locations. In addition, a number of extra points were digitized from the scalp, starting from the tip of the nose and moving on to the back of the head and also collaterally starting from the temple and moving ahead to the other side. These digitized coordinates of the shape of the head are used to reconstruct the ECD or MNE estimates if the MEG data are combined together with MRI data.

Before entering the shielded room the subject was instructed to touch the door frame to be "grounded" and asked to ensure that all metal objects are removed. Then the subject was seated to the MEG armchair and the HPI coil cables and the EOG cables were connected to the MEG system. The seat was lifted so that the subject's head was set within the helmet. The appropriate height of the seat was adjusted so that the vertex of the subject's head was touching the lining of the helmet so that the cortical distance to the SQUIDs was minimal. However, the head position and the stance of the subject was adjusted to be as comfortable as possible, because otherwise it would be problematic to stay still. Uncomfortable stance may cause excessive movements or tense the neck muscles which decreases the quality of the recording by adding artefacts to the signal.

### 3.4.2 Measuring data

During the MEG measurement the door of the shielded room was closed and the subject sat under a helmet-shaped measuring device in front of the screen and panel speakers. For safety reasons there was a small video camera and a microphone so the experimenter was able to observe and communicate with the subject.

The experiment started with sound intensity level adjustment where the detection level of the target tones were adjusted at 50% threshold level – to be "barely audible". The adjustment was made by using an *up-down procedure* (Levitt and others, 1971) where the signal-to-noise ratio was modified according to the responses of subject. The subject was instructed to respond with an optical device by a finger lift when they heard a target

tone. The target tones were 100-ms 1 kHz sinusoidal tones with 5 ms linear rise and fall times, which were presented at a random 2–8 second inter stimulus interval. SPL of the target tones was fixed at 65 dB, but level of the continuous white noise masker sound pressure was modified according to *hits* and *misses* of the subject. A hit ('+') was defined as a response preceding the onset of the stimuli within a time window of 0–1 s. Delayed answers (more than 1 s) or answers where no stimulus was presented were interpreted as a miss ('-'). The up-down procedure was applied according to the last 2–3 answers by increasing the noise level by 1 dB after a positive response sequence (++, +-+, -++) or decreasing the noise level by 1 dB after a negative response sequence (–, -+-, +–). The 50% threshold level was adjusted by several iterations so that the adjustment was not terminated after the first reversal, but continued until four reversals were obtained. The adjustment was done using an automatic presentation script so that the experimenter had time to simultaneously check the signal levels in the MEG channels and tag the faulty channels.

The psychophysical experiment was divided in seven different masking conditions: ±500 Hz, ±300 Hz, ±200 Hz, ±150 Hz, ±100 Hz, ±50 Hz and 0 Hz. The runtime of a single block (trial) was 8 minutes and the presentation order of trials was randomized across subjects (Figure 3.2, p. 55). However, one extra trial without a masker was presented in the beginning of each experiment as a control. The beginning and the end of a each trial was instructed on screen, as well as changes in the task conditions, which were taking turns in 2-minute periods. Thus a single trial contained 2 sub-blocks of the AttAud conditions and 2 sub-blocks of the AttVis conditions (Figure 3.2, p. 55). The starting order of the AttAud and AttVis conditions were also randomized among the trials.

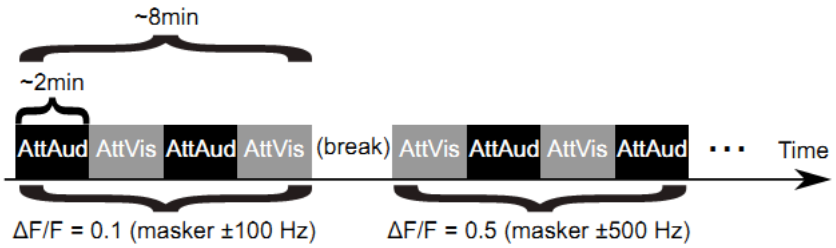



Figure 3.2: The experiment was divided in seven different masking conditions (±500Hz, ±300 Hz, ±200 Hz, ±150 Hz, ±100 Hz, ±50 Hz and 0 Hz) that were presented in a randomized order [adapted from Kauramäki et al., 2012].

During the trials the subjects were doing the behavioral task where they responded either to auditory or visual deviants by a finger lift. In the AttAud condition subjects tried to discriminate the frequency deviants, while they were watching a small fixation cross on the top of the computer screen[2]. In the AttVis condition the auditory stimulus setup was on the background, while subjects were detecting luminance pattern deviants (the Gabor patches) on the screen. In both conditions a '*hit*' was defined as a response to the target stimuli within a time window of 1.5–2.5 s, starting from the deviant appearance and

[2]Marked with tape.

ending to the presentation of the next stimuli. A '*false alarm*' was defined as an answer to a non-target stimuli or a delayed response – a response that was overlapping with next stimulus. A missing button press after target was defined as a '*miss*'.

Before the data acquisition subjects were asked to take a comfortable stance, where they could stay as immobile as possible until the end of the measurement block. The head position measurement was made at the beginning of each recording block by briefly energizing the head position indicator coils so that the coil locations could be localized in the MEG device coordinate system. After the HPI measurement the stimulus presentation and data acquisition were started. Subjects were also instructed to avoid muscular movements and eye blinks particularly right after the occurrence of a sound stimulus. There were small breaks between the blocks where subjects could rest and relax, blink and move their eyes freely. Typically one longer break ($\sim$ 5 min) was made somewhere in the middle of the experiment so that subjects could drink water in order to maintain alertness.

## 3.5 The data analysis

### 3.5.1 MEG data analysis

In the current study we used plain sensor data by taking vector sums from the gradiometer channel pairs:

$$G_p = \sqrt{\frac{\partial {B_z}^2}{\partial x} + \frac{\partial {B_z}^2}{\partial y}} \tag{3.1}$$

where the $\partial B_z/\partial x$ and $\partial B_z/\partial y$ are the two orthogonal field components.

The MEG data contained initially measurements from 18 subjects. However, as mentioned previously, four subjects were excluded from the data analysis, because appropriate number of response segments (epochs) could not be collected. Faulty epochs were eliminated with the rejection criteria, which was triggered by excess in any of the following parameters: 3000 fT/cm in gradiometer channel, 400 fT in magnetometer channel or 150 $\mu$V in EOG channel. The reasons for disqualifications were claustrophobia, signal contamination by strong artefacts (e.g. hair dye including metal), excessive number of eyeblinks and extremely low SNR.

Thus data sets from 14 subjects were qualified to the final MEG data analysis. Each individual data set contained at least 100 accepted epochs of each stimulus type (standards and deviants) and condition type (AttAud and AttVis). Typically the number of epochs in a single data set was between 100 and 120. The epochs were time-locked to the stimulus onset, 200 ms before the onset and 800 ms after the onset. The stimulus and the corresponding brain response were synchronized by using the log-files of Presentation software and the microphone channel, which picked up the onsets of auditory stimuli.

The response segments were grouped per subject according to the eight masker types (±500Hz, ±300 Hz, ±200 Hz, ±150 Hz, ±100 Hz, ±50 Hz, 0 Hz and no-masker). Inside each group the response segments were separated further according to the stimulus type and the condition type, so that the grand means could be calculated among subjects. Before this broken MEG channels were disabled and the healthy channels were low-pass-filtered with a 40-Hz cut-off frequency and corrected to a 200-ms prestimulus baseline. The grouped responses were offline-averaged by using MNE suite (v2.7.0, http://www.nmr.mgh.harvard.edu/martinos/userInfo/data/sofMNE.php) command-line scripts, to get the grand averages of AEFs.

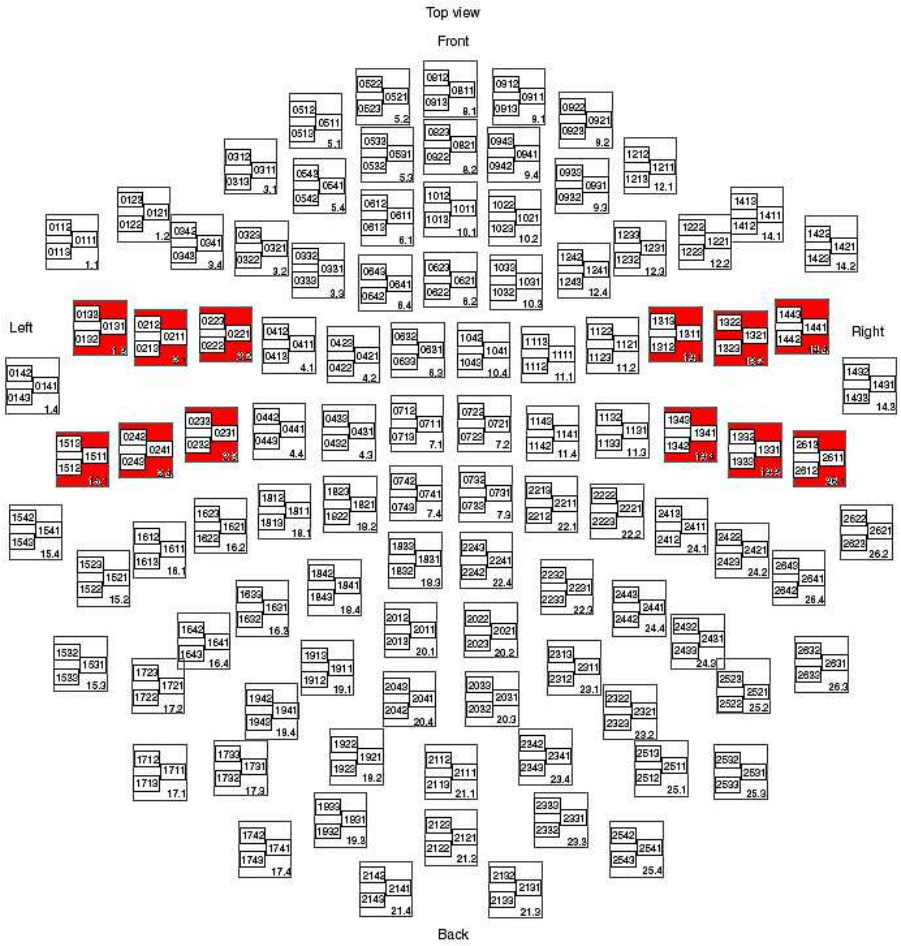



Figure 3.3: The selected triplets (planar gradiometer pair + magnetometer) are marked with red color. The helmet shaped sensory array is flattened into a plane.

The gradients of auditory evoked fields (*gradient AEFs*) were calculated separately for left and right hemisphere from a fixed subset of 6 planar gradiometer pairs (Figure 3.3, p. 57). In other words, singular gradient responses were used to represent the activities in left and right auditory cortices. The channel distributions were selected by using spherical symmetry so that the channel subsets represented similar auditory areas from left and right hemispheres. The channel selection was made by observing the robustness of appropriate channels with visualization softwares, such as Xplotter (includes in Elekta Neuromag

software) and mne_browse_raw (including in MNE Suite). The gradients were computed with Matlab (R14, Math-Works Inc., Natick, MA, USA) alongside with MNE Matlab Toolbox (Version 2.7.0, http://www.martinos.org/mne/manual/matlab.html), so that first the vector sums of each gradiometer channel pairs were calculated and then the averages were taken from the remaining 6 vector sums to obtain the gradient AEFs.

Latencies and amplitudes of the onset response peaks were obtained directly from the calculated gradient amplitude by using a Matlab function that detects latency of the maximum amplitude within a limited time scale (90 – 220 ms). The N100m peak amplitudes was averaged by using $\pm$10 ms time window around the absolute peaks. Sustained fields were obtained similarly, but because of the response durability, the amplitudes were averaged by four 100-ms time intervals. The sustained fields were determined with respect to N100m component by using four fixed time windows in latency range of 100–200 ms, 200–300 ms, 300–400 ms and 400–500 ms after the N100m peak latency. The amplitudes of each latency range – LR01, LR02, LR03 and LR04 – were averaged within group's time window (*e.g.* LR01 within 100–200 ms).

The grand means were computed from individual gradient AEFs by averaging responses among all subjects (N=14 ). The peak amplitudes and latencies were obtained similarly than in the case of the individual gradient AEFs. The N100m peak amplitudes was averaged by using $\pm$10 ms time window around the absolute peaks and the SF amplitudes in four 100-ms time intervals – LR01, LR02, LR03 and LR04.

Finally, the grand mean AEFs and individual AEFs were mapped to each permutation – stimulus type (standards versus deviants) and condition type (AttAud versus AttVis). However, only standard evoked auditory N100m and SF components were used in the analysis, since neural responses evoked by deviants (e.g. mismatch field; MMF) reflect neural processing relating to sudden changes in auditory environment, which do not include in the scope of current study. Thus, the standard evoked auditory N100m and SF components (the neural data) were mapped to the notch width of the noise maskers (Function 2.14, p. 38) – and were comparable between task conditions AttAud and AttVis. The corepoint was to quantify how the parametric changes in the level of auditory attention (P1) and background noise (P2) reflect to the processing at auditory cortices.

### 3.5.2 Behavioral data analysis

The behavioral data were used in parallel with MEG data to investigate the relation between auditory cortex activity and auditory performance. The auditory and visual task responses: hits, false alarms, misses and reaction times, were extracted from the log-files of Presentation software. Individual *hit rates* (HR), *false alarms* (FA), *misses* and *reaction times* (RT) were determined to each data sets ($\pm$500Hz, $\pm$300 Hz, $\pm$200 Hz, $\pm$150 Hz, $\pm$100 Hz, $\pm$50 Hz, 0 Hz and no-masker) and conditions (AttAud and AttVis). Because the HR value did not take subjects strategy to answer (response bias) into account, the *d-prime* (d') values ware calculated also. D-prime is a measure of detection sensitivity,

which is the difference between the signal (HR) and the signal+noise (FA) in Z-domain[3]. The detection sensitivity can be calculated from Equation:

$$d' = z(HR) - z(FA) \tag{3.2}$$

where The output of the equation is 0 when HR = FA and starts to increase in tandem with HR/FA. The highest possible d' is 6.93 and typical values lie somewhere between 1 and 3.

The grand means of hit rates, d-primes, false alarms and misses were computed from the individual behavioral data (N=14).

### 3.5.3 Statistical analysis

The results of the MEG study were statistically analyzed by *Repeated-measures* (RM) *analysis of variance* (ANOVA). This method is appropriate in experiments where the same subject(s) is measured several times. The *dependent variables* – peak amplitudes of N100m and SF components – were analyzed using *three-way* RM ANOVA. The following three *independent factors*: MASKER TYPE (notch width), CONDITION (AttAud/AttVis) and HEMISPHERE (left/right) were used. Also the N100m latencies were tested with the three-way repeated-measures ANOVA with similar factors. Following interaction terms MASKER TYPE $\times$ CONDITION, MASKER TYPE $\times$ HEMISPHERE and CONDITION $\times$ HEMISPHERE were also analyzed in the three-way RM ANOVA. The three-way RM ANOVA cases were computed with Matlab (R14, Math-Works Inc., Natick, MA, USA).

The behavioral data were analyzed by *two-way* RM ANOVA where dependent variables (d', HR, RT) were compared to factors: MASKER TYPE and CONDITION. There were missing values in reaction times table because all subjects did not detect deviants with thin notches ($\pm$50 Hz and 0 Hz). The missing RT values were replaced with *expectation-maximization* (EM) algorithm (Dempster et al., 1977), which is an iterative method for finding maximum likelihood or *maximum a posteriori* (MAP) estimates for unknown parameters. The two-way RM ANOVA cases were computed with R (http://www.R-project.org) version 2.14.0.

[3]Z-transform converts a discrete time-domain signal into a complex frequency-domain representation.

# Chapter 4

# Results

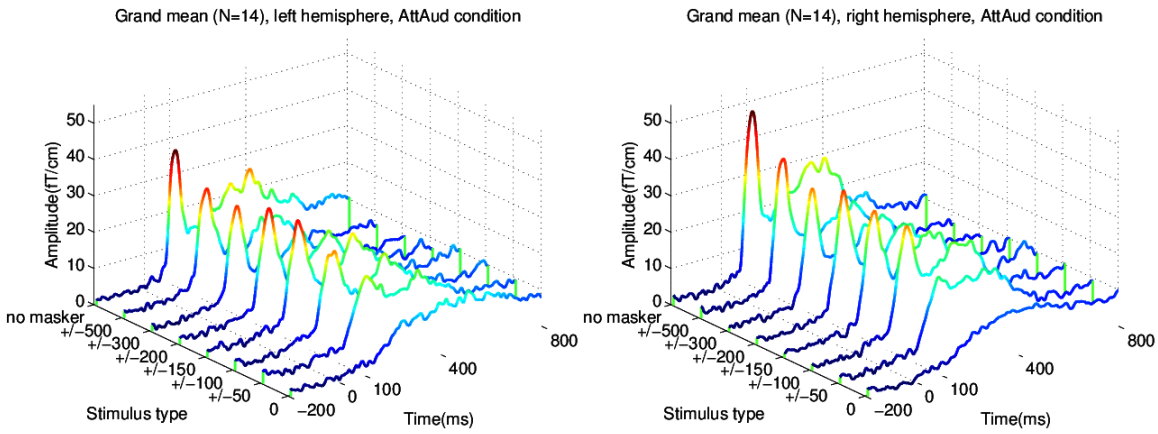
## 4.1 Results of the MEG study

The auditory evoked fields from both left and right hemispheres are shown in the grand averaged plots (Figure 4.1, p. 61), which were averaged across all 14 subjects. Individual waveforms (7+1) in the figure represents different masking conditions. Transient N100m components (100–200 ms post-onset) are the first peaks in the waveforms and the sustained field components are the subsequent flat peaks lasting several hundred ms – approximately 200–600 ms post-onset.

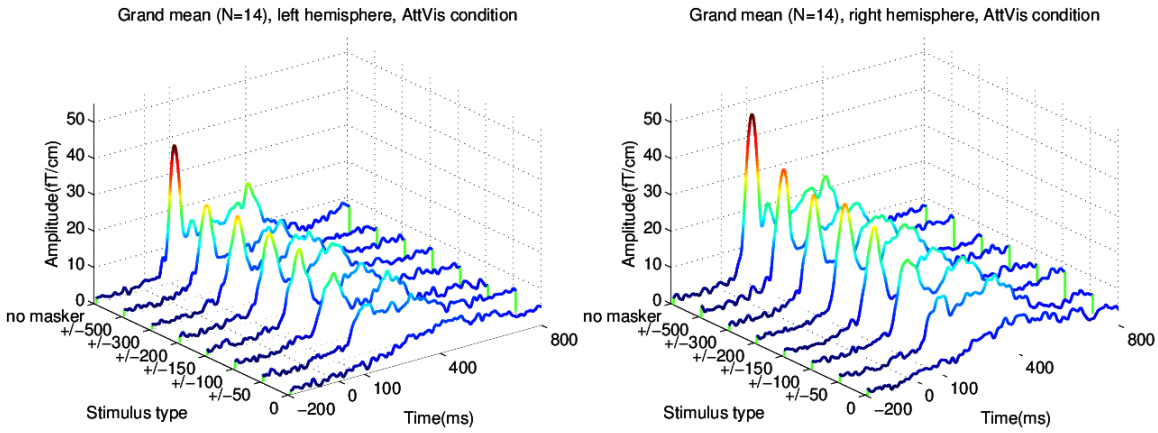
### 4.1.1 N100m peak amplitudes and latencies

The peak amplitudes of N100m components are shown in Figure 4.2 (p. 61) and their latencies in Figure 4.3 (p. 62). The amplitude and latency differences between subjects are presented with error bars indicating the standard error of the mean (SEM). A clear trend relating to concurrent noise was that narrowing the notch of the noise masker decreased the N100m peak amplitudes ($F(6,78) = 6.778$; $p < 0.0001$) (Figure 4.2, p. 61) and delayed the latencies of the N100m peaks ($F(6,78) = 61.871$; $p < 0.0001$) (Figure 4.3, p. 62). The peak amplitudes decreased from 36.27 fT/cm to 13.90 fT/cm and latencies increased from 125 ms to 210 ms, when the notch bandwidth was narrowed from $\pm$500 Hz to 0 Hz. In addition, the amplitudes were slightly larger in the right hemisphere ($F(1,13) = 7.650$; $p = 0.0069$). In AttVis control condition the amplitudes were 18% stronger in the right hemisphere (29.38 fT/cm vs. 24.05 fT/cm) and in AttAud condition 11% stronger (33.91 fT/cm vs. 30.33 fT/cm).

A clear trend relating to focus of attention was that the N100m peak amplitudes were stronger during the AttAud tasks than during the AttVis tasks (Figure 4.2, p. 61). The amplitude increment under auditory attention was bihemispheric and statistically significant ($F(1,13) = 81.973$; $p < 0.0001$). However, the interaction between the masker type and the condition was not statistically significant ($F(6,78) = 1.183$; $p = 0.3224$). Also

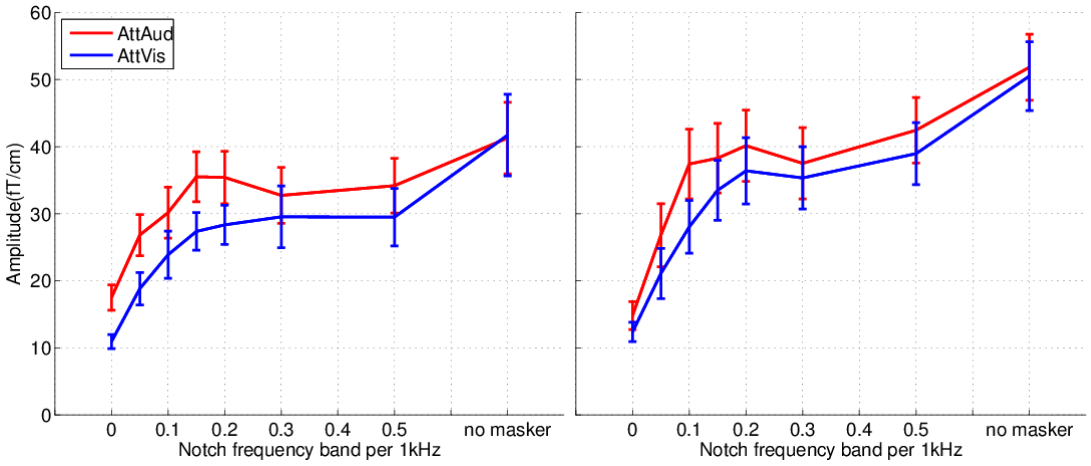



(a) AttAud condition



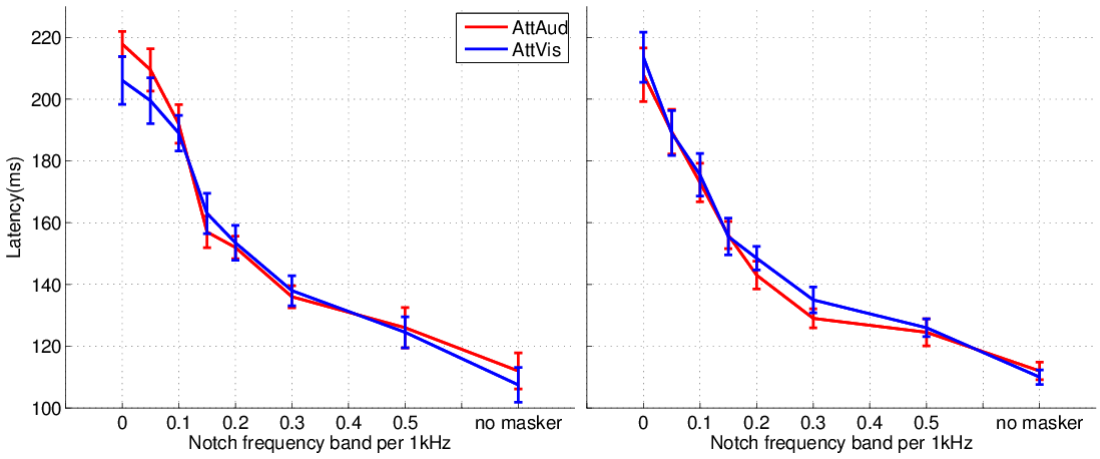

(b) Attvis condition

Figure 4.1: Auditory evoked fields measured under different notched-noise maskers.



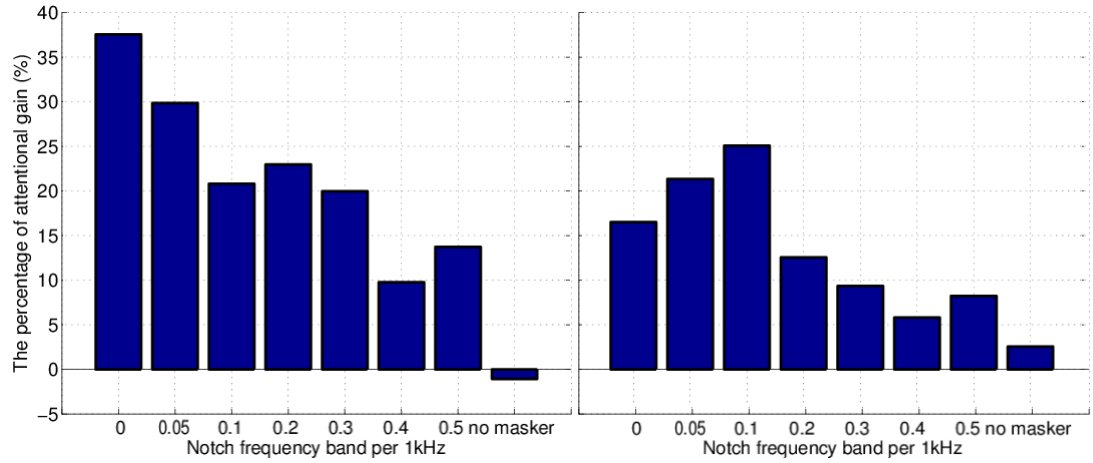

(a) Left hemisphere (b) Right hemisphere

Figure 4.2: N100m peak amplitudes (±SEM).




(a) Left hemisphere (b) Right hemisphere

Figure 4.3: N100m latencies (±SEM).




(a) Left hemisphere (b) Right hemisphere

Figure 4.4: The percentual gain of N100m amplitudes (AttAud > Attvis) under different notched-noise maskers.

the latencies of N100m peaks were quite similar between AttAud and AttVis conditions (Figure 4.3, p. 62), and thus no statistical significance could be pointed out ($F(1,13) = 0.000$; $p = 0.9914$).

The Figure 4.4 (p. 62) shows the percentages[1] of amplitude increment according to the masker type with respect to auditory and visual data sets. The gain effect of auditory attention was most salient in noise conditions with narrow notch widths (±200 Hz, ±150 Hz, ±100 Hz, ±50 Hz and 0 Hz). The gain increment was 33% (29.08 fT/cm vs. 21.86 fT/cm) on the left hemisphere and 20% (31.48 fT/cm vs. 26.27 fT/cm) on the right hemisphere, when the percentages were calculated from averages of four notch types: ±200

[1] (AttAud-Attvis)/AttVis*100

Hz, ±150 Hz, ±100, ±50 Hz Hz and 0 Hz. Similar percentages of all seven masker types were 26% (30.33 fT/cm vs. 24.05 fT/cm) on left and 15% (33.91 fT/cm vs. 29.38) on right. The gain increment was also slightly stronger in left auditory cortex with notch widths narrower than ±200 Hz (CONDITION X HEMISPHERE: $F(1,13) = 6.802$; $p = 0.0106$).

### 4.1.2 SF amplitudes and latencies

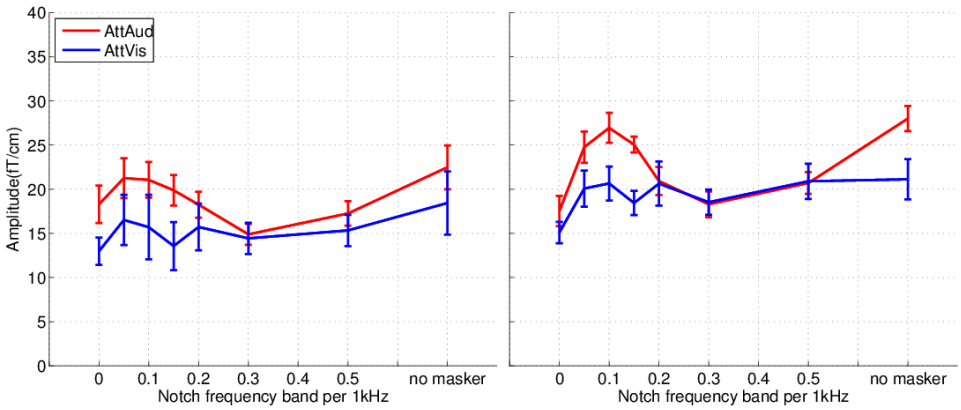
The SF amplitude suppression as a function of the masker type is presented in Figure 4.5 (p. 64). Similarly to the transient responses, the masker type affected significantly to the SF component amplitudes. During the visual control condition SF component amplitudes were stronger in the right hemisphere than in the left hemisphere. The difference was 19% in LR01, 15% in LR02, 8% in LR03 and 2% in LR04. This hemispheric asymmetry was significant in the first two time frames LR01 ($F(1,13) = 34.880$; $p < 0.0001$) and LR02 ($F(1,13) = 15.372$; $p = 0.0002$), but insignificant in LR03 ($F(1,13) = 2.551$; $p = 0.1137$) and LR04 ($F(1,13) = 0.339$; $p = 0.5620$).

The focus of attention affected also to the sustained fields so that the SF component amplitudes were stronger during the AttAud tasks than during the AttVis tasks (Figure 4.5 p. 64). The amplitude increment was bihemispheric and statistically significant in all latency ranges: LR01 ($F(1,13) = 48.471$; $p < 0.0001$), LR02 ($F(1,13) = 53.042$; $p < 0.0001$), LR03 ($F(1,13) = 54.855$; $p < 0.0001$) and LR04 ($F(1,13) = 34.477$; $p < 0.0001$). In addition, MASKER TYPE × CONDITION interaction was significant at 300–600 ms post-onset: LR01 ($F(6,78) = 3.912$; $p = 0.0016$), LR02 ($F(6,78) = 2.777$; $p = 0.0159$) and LR03 ($F(6,78) = 2.776$; $p = 0.0159$). The last time frame LR04 ($F(6,78) = 0.915$; $p = 0.4877$) was statistically insignificant. Moreover, an interesting anomaly in the curves was at ±300 Hz notch (Figure 4.5, p. 64), where the gain effect was zero or even below.

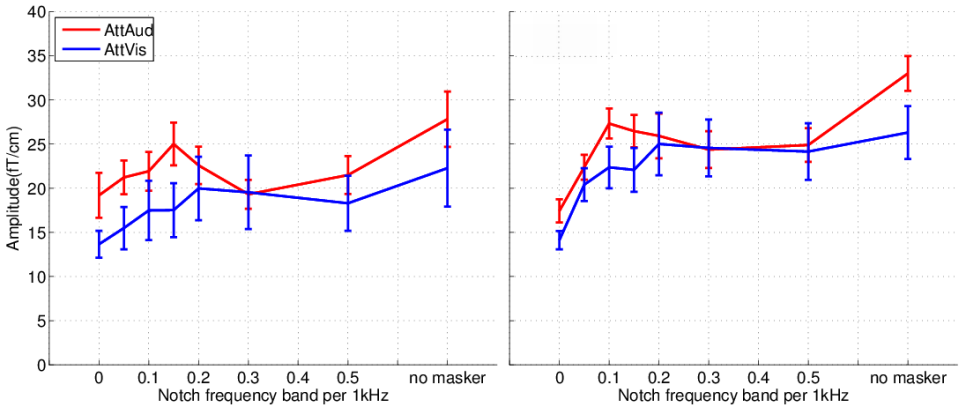
In the case of sustained fields, the amplitude increment relating to the auditory attention was also more powerful in the left hemisphere (Figure 4.6 p. 65). The percentages of increase – the mean value of all seven masker types – were 26% (LR01: 18.70 fT/cm vs. 14.89 fT/cm), 24% (LR02: 21.53 fT/cm vs. 17.42 fT/cm), 27% (LR03: 16.75 fT/cm vs. 13.15 fT/cm) and 23% (LR04: 14.41 fT/cm vs. 11.72 fT/cm) on the left hemisphere. On the right hemisphere the corresponding percentages were 15% (LR01: 22.03 fT/cm vs. 19.19 fT/cm), 11% (LR02: 24.11 fT/cm vs. 21.81 fT/cm), 17% (LR03: 17.56 fT/cm vs. 15.00 fT/cm) and 21% (LR04: 14.59 fT/cm vs. 12.03 fT/cm).

The lateralization of amplitude increment was similarly – like in case of the N100m responses – most salient in noise conditions with narrow notch widths (±150 Hz, ±100 Hz, ±50 Hz and 0 Hz). The percentages of amplitude increment[2] on the left hemisphere were 34% (LR01: 20.20 fT/cm vs. 15.06 fT/cm), 34% (LR02: 20.78 fT/cm vs. 15.54 fT/cm), 41% (LR03: 17.29 fT/cm vs. 12.25 fT/cm) and 35% (LR04: 15.23 fT/cm vs. 15.54 fT/cm), and on the right hemisphere were 24% (LR01: 23.08 fT/cm vs. 18.61 fT/cm), 18% (LR02: 22.38 fT/cm vs. 18.96 fT/cm), 26% (LR03: 16.60 fT/cm vs. 13.13 fT/cm) and 29% (LR04: 14.51 fT/cm vs. 11.22 fT/cm).

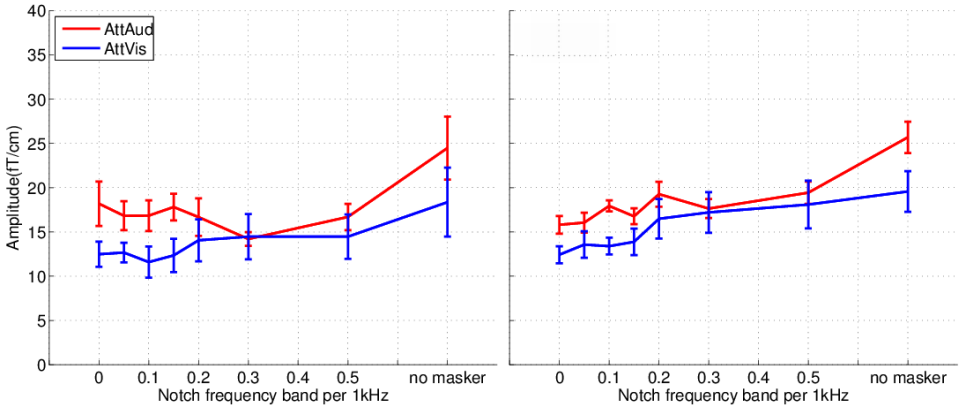
[2]The mean value of masker types: ±150 Hz, ±100 Hz, ±50 Hz and 0 Hz.



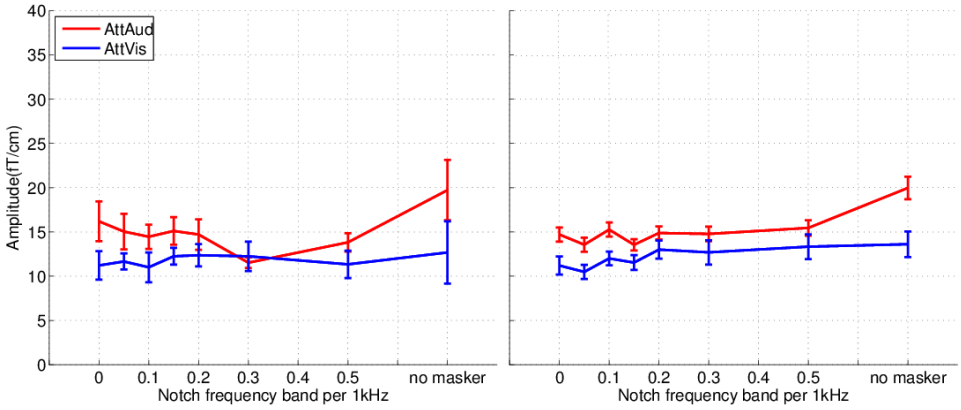

(a) Latency range 1; 100–200 ms after the N100m




(b) Latency range 2; 200–300 ms after the N100m

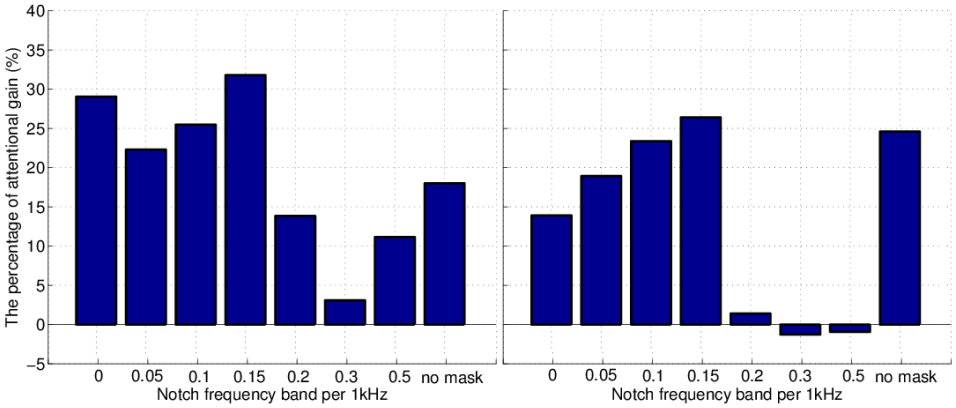



(c) Latency range 3; 300–400 ms after the N100m

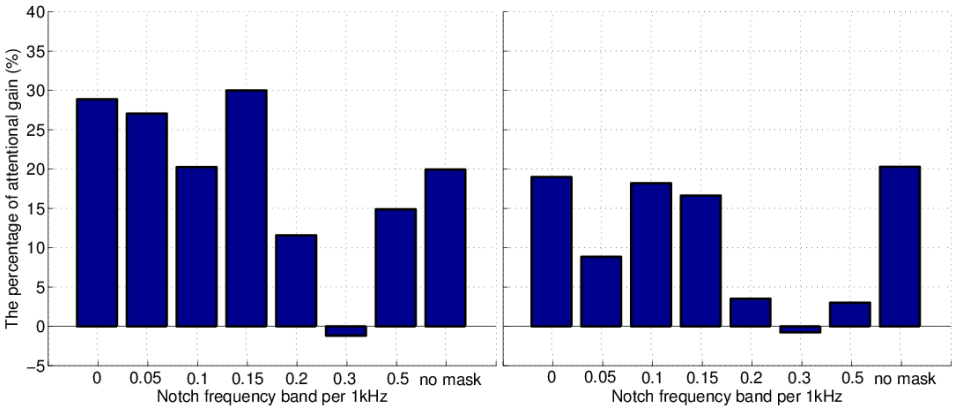



(d) Latency range 4; 400–500 ms after the N100m

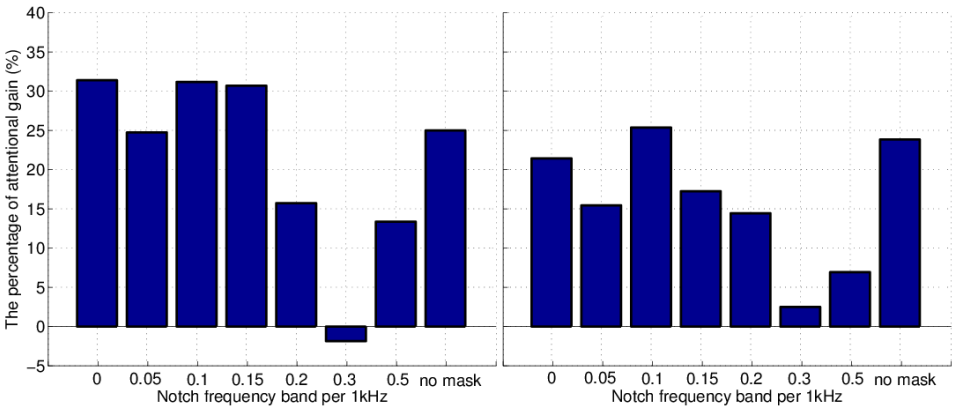
Figure 4.5: Auditory evoked SF fields (±SEM) measured under different notched-noise maskers. Left side plots represent responses from the left hemisphere and right side plots responses from the right hemisphere.



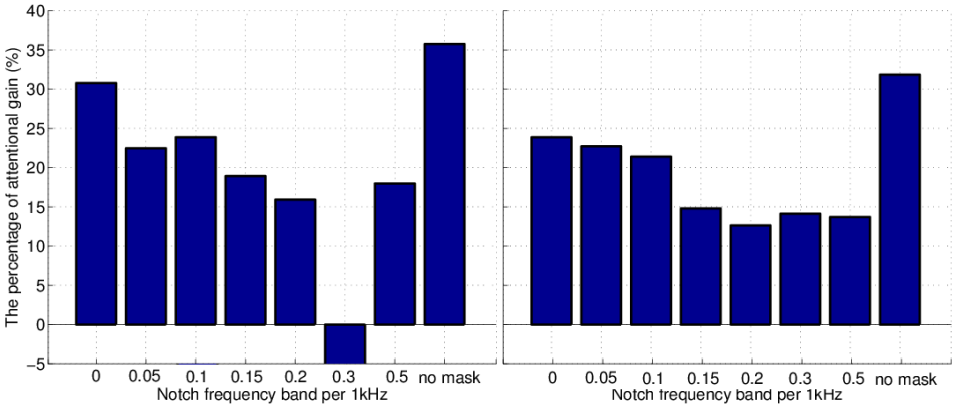

(a) Latency range 1; 100–200 ms after the N100m




(b) Latency range 2; 200–300 ms after the N100m




(c) Latency range 3; 300–400 ms after the N100m




(d) Latency range 4; 400–500 ms after the N100m

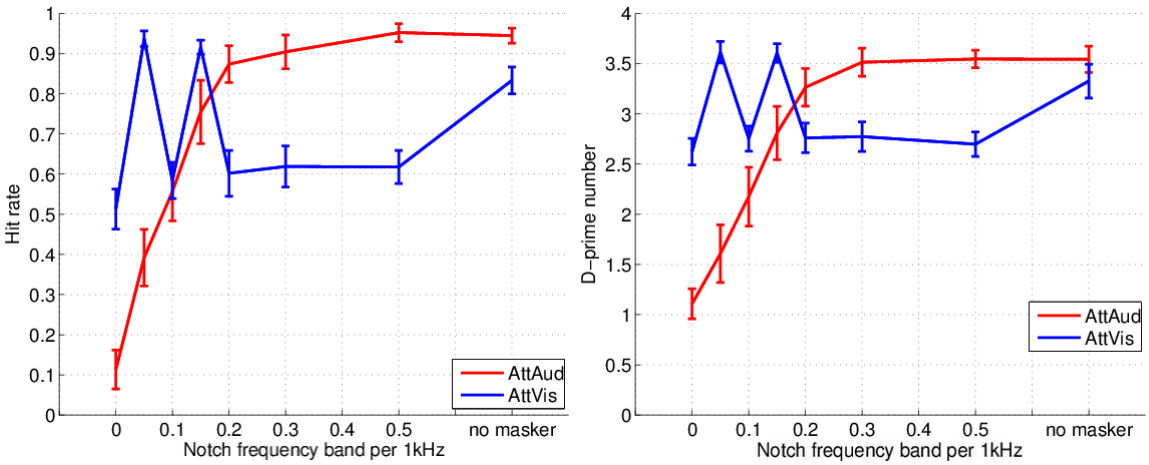
Figure 4.6: The percentual gain of SF amplitudes (AttAud > Attvis) under different notched-noise maskers. Left side plots represent the left hemispheric gain and right side plots the right hemispheric gain.

## 4.2 Results of the behavioral test

The auditory and visual task performances are presented in Figure 4.7a (p. 67) and 4.7b (p. 67) as a function of masker type. 4.9% (11/224) of the reaction time data were replaced by EM estimates (Dempster et al., 1977) because all subjects could not detect auditory stimuli in the most difficult noise conditions (±100 Hz, ±50 Hz, 0 Hz). However, the SEM values were calculated with the original data. In addition, some indicative plots of auditory performance and N100m components correlation presented in Figure 4.9 (p. 68) and 4.10 (p. 68)

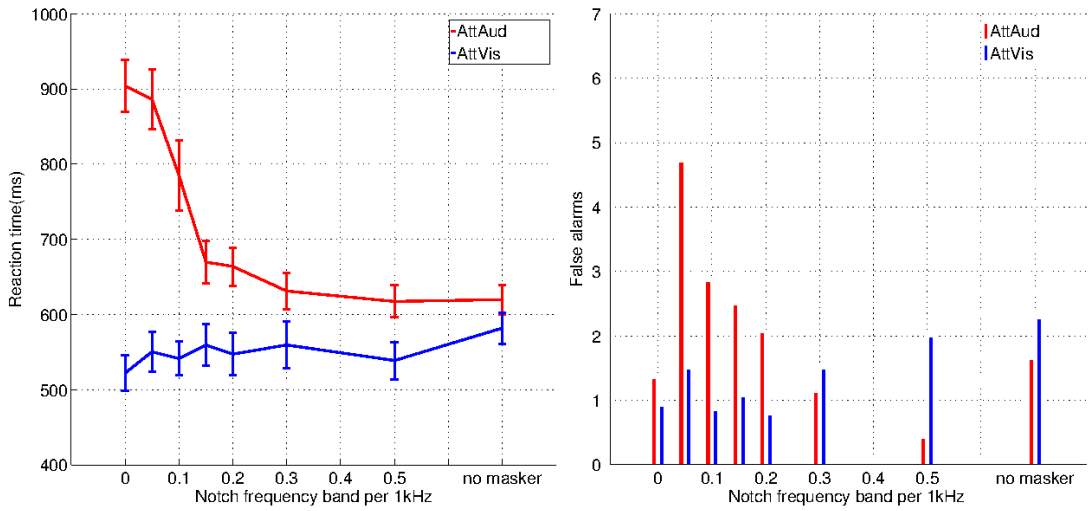
The masker type effected of significantly to both, detection sensitivity and reaction times (D-prime: $F(6,78) = 15.58$; $p < 0.0001$ and RT: $F(6,78) = 7.45$; $p < 0.0001$). In addition, the CONDITION X BEHAVIORAL PERFORMANCE (D-prime: $F(6,78) = 18.04$; $p = 0.000344$ and RT: $F(6,78) = 145.45$; $p < 0.0001$) and MASKER TYPE X CONDITION were statistically significant (D-prime: $F(6,78) = 20.63$; $p < 0.0001$ and RT: $F(6,78) = 9.40$; $p < 0.0001$). In AttAud condition the d'-curve was first almost flat between ±500 Hz and ±200 Hz, but started to decline steeply when the notch width reached ±200 Hz (Figure 4.7, p. 67). In AttVis condition the shape of the curve was relatively flat except the exceptional "sawtooth" pattern in the curve between ±200 Hz and 0 Hz. The dependancy of masker width and shape of the d'-curve was different between conditions, since the noise did not effect straightforward to the visual task performance. However, the "sawtooth" pattern in the curve indicates that there exists an indirect causation.

The forms of d'-plots resemble the N100m amplitude plots (AttAud; left and right). Figure 4.9 illustrates the correletion (p. 68), so that the y-axes (d' and N100m magnitudes) were scaled-to-fit (atrbitrary units). The correletion of reaction times (AttAud) and N100m latencies (AttAud; left and right) is illustrated in Figure 4.10 (p. 68); y-axes of (RT and N100m latencies) were also scaled-to-fit (atrbitrary units).



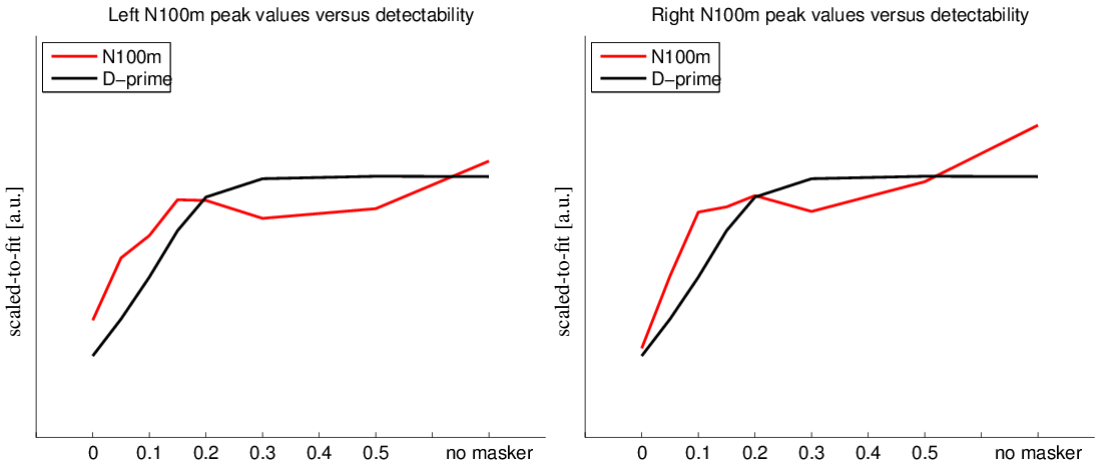

(a) Hit rates (b) Detectability

Figure 4.7: Hit rates and detectability (±SEM)



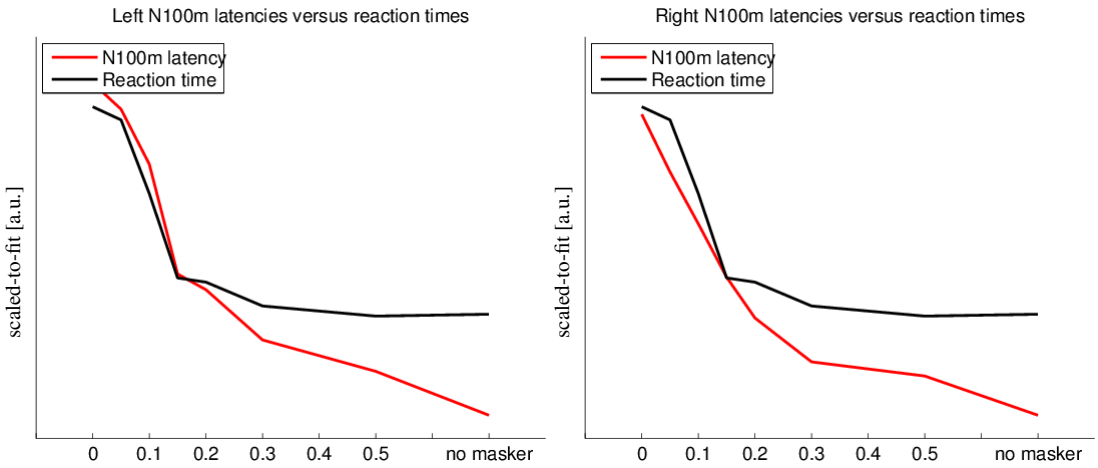

(a) Reaction times (b) False alarms

Figure 4.8: Reaction times (±SEM) and false alarms




(a) Left N100m vs. d’ (b) Right N100m vs. d’

Figure 4.9: N100m peak amplitudes versus detectability




(a) Left N100m latency vs. reaction time (b) Right N100m latency vs. reaction time

Figure 4.10: N100m latencies versus reaction times

# Chapter 5

# Summary and discussion

The human auditory system can be compared to a spectrum analyzer. Complex sound patterns are decomposed to their frequency components in the inner ear's cochlea, and this spectral information is then encoded to the tonotopic structures of the auditory nervous system. The frequency resolution of hearing (i.e. the frequency selectivity) is limited by the mechanical properties of the cochlea, and further by the tonotopy of auditory cortex. When two separate tones are close enough in frequency – within the same critical bandwidth – they tend to stimulate identical receptive fields at auditory cortex and cause perceptual interference called auditory masking. However, under difficult noise conditions we can select relevant auditory information even if the signal-to-masker ratio (or SNR) seems to be too poor for frequency discrimination. The prevailing neurophysiological studies presume that the selective auditory attention facilitates the cortical processing of task-relevant inputs (gain model: Hillyard et al. (1973); Rif et al. (1991); Woldorff et al. (1993)) and sharpens the selection of relevant inputs by neural plasticity (tuning model: Ahveninen et al. (2006, 2011); Altmann et al. (2008)).

The current study aimed to investigate the attention related short-term neural plasticity changes in the human auditory cortex that will push the physical limits of the frequency resolution during concurrent noise. The experimental paradigm took advantage of the spatio-temporal accuracy of the modern functional brain imaging technology MEG and combined it to the theory of auditory filters. In the auditory filter theory the function of the tonotopic auditory cortex is modeled with a filter bank of band-pass filters (Fletcher, 1940; Patterson, 1986; Glasberg and Moore, 1990), whose pass-band is an approximation of the critical band at a fixed center frequency. In the psychophysical experiment the neural and behavioral correlates of these auditory filters were measured by a notched-noise experiment (see Section 2.4.3, p. 37) where a band-stop filtered spectral "notch" was narrowed gradually around the target signal. Narrowing the notch of the masker increases the auditory masking effect, which causes a suppressive effect to the tone-evoked response magnitudes and lowers the signal detection. The essence of the study was to investigate, how selective attention biases the masker-induced suppression curves – adjusts the shapes of the auditory filters. In the functional model the increase in filter's height can be interpreted as the gain factor and the decrease in filter's passband as the sharpening

factor.

In the MEG measurements N100m onset fields (latency $\sim$ 100 ms) and the sustained fields (SF; latency $\sim$ 200–600 ms) were evoked by 1 kHz pure tones that were embedded within the continuous-time notch-filtered noise maskers. We used seven different masker types between $\pm$500 Hz and 0 Hz (see Section 3.2, p. 55). The neural responses were measured with a whole-head MEG device during a behavioral task, where subjects (N=14) identified 1020-Hz target tones (P=0.1) from the more frequently occurring 1000-Hz standard tones, or respectively made a visual control task where identical auditory stimulus setup was playing on the background. Both N100m and SF components were calculated separately for the left and the right hemisphere from a fixed subset of 6 planar gradiometer pairs, which were selected mirror-symmetrically over the auditory cortices. The amplitudes of standard evoked auditory N100m and the SF components[1] were mapped to the notch width of the noise maskers. These masker-induced suppression curves (Function 2.14, p. 38) were comparable between auditory and visual data sets, since the neural responses were evoked with identical auditory stimulus setup. The behavioral measures (d' and RT) were used in parallel with the MEG data to investigate the relation between the auditory cortex activity and the auditory performance.

Our experimental data showed that selective attention has influence on both auditory evoked N100m and SF components (Figure 4.1, p. 61). The most robust attentional enhancements were observed in difficult noise conditions, where the bandstop filtered notches were within the critical band at 200–500 ms post-onset. In addition, the effect was more pronounced in the left auditory cortex with the thinnest notches; the impact strengthened progressively when the edges of white noise were shifted closer to the target frequency.

Our results indicate that the neural mechanism of frequency-specific attention is based on bihemspheric neural gain when the masker notches are moderate and the discrimination conditions are pretty easy. In the most demanding noise conditions the contribution of the left auditory cortex becomes stronger, while the segregation of relevant sounds from noise requires very sharp filtration. The amplitude increment with the thinnest notches could also relate to the increase of the inhibitory that is needed at auditory cortex to reduce the stimulation of noise.

## 5.1 Effects of notched-noise masker

The tone-evoked N100m and SF response patterns arise mainly from the gyri of the folded auditory cortices (Pantev et al., 1995; Bidet-Caulet et al., 2007) where the pyramidal cells are optimally organized to genarate post-synaptic potentials that sum up to macroscopic ERF and ERP signals (Hämäläinen et al., 1993). The amplitudes and latencies of the

[1] Only the standard evoked auditory N100m and SF components were used in data analysis, since these components are modulated by top-down controlled attention, whereas deviant evoked fields are modulated by bottom-up processes (Knudsen, 2007).

N100m onset responses are dependent on the spectral and temporal features of the preceding auditory events.

In the current notched-noise experiment the impact of spectral masking was studied by masking a sequence of auditory stimuli (300-ms pure tones) with the continuous-time notched-noise maskers (Figure 3.1, p. 52). As it was expected (hypothesis a, p. 47), narrowing the notch width of the notched-noise masker (±500 Hz – 0 Hz) impaired the pitch discrimination sensitivity (d') and suppressed the amplitudes of the auditory evoked N100m components. The masker-induced suppression effect was in line with the previous notched-noise studies (*e.g.* Hari and Mäkelä (1988); Sams and Salmelin (1994)), where the amplitudes of tone-evoked responses degenerated, while the edges of white noise were shifted closer to the target frequency. As a novel finding – documented in MEG study of Kauramäki et al. (2012) – the masker-induced suppression was statistically significant also in SF components with 300–400 ms post-onset.

The measurements confirmed also the hypothesis b (p. 48); the masker-induced suppression was minor with wide notch widths (±500 Hz and ±300 Hz), but the supression strengthened dramatically when the masker notches were narrowed within the critical band. The critical band is approximately ±160 Hz at 1 kHz (Fletcher, 1940; Zwicker and Terhardt, 1980). On the right side the pronounced decay started slightly earlier (±200 Hz notch) than on the left side (±150 Hz notch). In addition, the amplitude decay was slightly steeper on the right side (Figure 4.2, p. 61, Blue line). The narrowing the notch of the masker effected also to the N100m latencies, which were delayed linearly from 125 ms to 210 ms (Figure 4.3, p. 62) when the notch width was decreased from ±500 Hz to 0 Hz.

In the frequency discrimination task the masking started to impair the detection of deviants prominently beyond ±200 Hz notches. With ±200 Hz notch the hit-rate was 0.87 and thereafter 0.75, 0.56, 0.39 and 0.11 when the notch was narrowed in ±50 Hz steps. The detection sensitivity values in similar cases were 3.26, 2.81, 2.18, 1.61 and 1.11. With 0 Hz notches the deviants were masked with pure white noise and the detection rate of deviants dropped near 10%. The task difficulty in such a case was observable also in N100m components. The reaction times increased linearly when the notches were narrowed, which relates to the rise of difficulty level.

The masker-induced suppression relates closely to the tonotopic nature of the human auditory system. The analysis of frequency information starts already at the cochlea, which encodes the spectral properties of the auditory stimulus (pressure waves) to the higher stages of auditory system (Ashmore, 2008). The tonotopic structures have been located in the primary auditory cortex and the belt areas (Talavage et al., 2004; Humphries et al., 2010). Masking frequencies that overlap at the basilar membrane with targets overlap also at auditory cortex where neural populations are activated repeatedly by similar sound frequencies. The constant stimulation of wideband noise causes synaptic depression that reduces the peak amplitudes of the onset responses (Budd et al., 1998; Rosburg et al., 2004).

## 5.2 Effects of selective auditory attention

Typically the attention-related amplitude enhancements on auditory N100(m) (Hillyard et al., 1973; Woldorff and Hillyard, 1991; Rif et al., 1991; Woldorff et al., 1993) and SF components (Picton et al., 1978; Arthur et al., 1991) have been explained by a multiplicative effect where the task-relevant inputs are facilitated and the non-relevant inputs are suppressed at auditory cortices (Hillyard et al., 1998). However, the gain model is not adequate enough in notched-noise experiments where the notches are narrowed below critical band; to the limits of frequency resolution where the gain mechanism would amplify also the non-relevant inputs.

The recent notched-noise paradigm studies (Kauramäki et al., 2007; Okamoto et al., 2007b) have found some evidence of the selectivity increase that would work in tandem with neural gain (tuning model). fMRI and EEG/MEG studies (Ahveninen et al., 2006; Altmann et al., 2008; Ahveninen et al., 2011) suggested that the neural mechanism (selectivity increase) can be based on neural adaptation; short-term plasticity changes in the secondary auditory areas. Jääskeläinen et al. (2007) propose that the sharpening can be implemented neurally by lateral inhibition (i.e. center-excitation and surround inhibition), which can reduce the overlap between neural populations that represent different frequencies.

In the present study we observed larger standard evoked auditory N100m (Figure 4.2, p. 61) and sustained field strengths (Figure 4.5, p. 64) under selective auditory attention as contrasted to the visual condition. As was expected in the hypothesis c (p. 48), the modulatory effect of selective attention was more powerful in narrow notch conditions, when the masker notches were within the critical band ($\leq \pm 160$ Hz). Moreover, interestingly the effect was stronger in the left hemisphere. The attention-related gain on SF amplitudes was approximately 38% on the left hemisphere and 23% on the right hemisphere with the four thinnest notches ($\pm 150$ Hz – 0 Hz; Figure 4.6, p. 65). A similar but slightly weaker lateralization was also observable on N100m components (Figure 4.4 p. 62) with the narrowest notches ($\pm 50$ Hz – 0 Hz). Also the behavioral data support the selectivity increase; the hit-rates in the frequency discrimination task did not drop to zero even in the most demanding noise conditions[2].

The current findings match well with the fact that complex linguistic functions, such as speech perception, are lateralized to the left auditory cortex (Liegeois-Chauvel et al., 1999). Speech signals are actually rapidly changing broadband sounds, so it is quite natural that the left auditory cortex is more active during the most difficult frequency discrimination conditions. By contrast, in our measurements the attentional gain was bihemispheric and even stronger in the right hemisphere in conditions where the notches were clearly wider than the critical band ($\pm 500$ Hz and $\pm 300$ Hz). The right-hemispheric lateralization has been found in experiments where no noise was on the background (Kanno et al., 1996) and experiments where attentional load was low (Alcaini et al., 1995). Thus, the crucial points of the left-lateralized effect seems to be that the cognitive load and the

[2]With $\pm 150$ Hz notch the HR was 0.75, with $\pm 100$ Hz notch 0.56, with $\pm 50$ Hz notch 0.39 and 0 Hz notch 0.11.

auditory masking levels are high. The left-hemispheric dominance has been found previously in a notched-noise study (Okamoto et al., 2007a) on transient N100m components, but in our measurements the asymmetry was stronger in later latencies at 200 – 500 ms post-onset. Moreover, the left side SF amplitudes correlated well with the auditory task performance [3].

The neural and behavioral results of the current study support the neural tuning model. The gain based bilateral effects were visible in moderate noise conditions at 100–200 ms post-onset and more robust left-hemispheric effects in the most demanding noise conditions at 200–500 ms post-onset. The results suggest that the neurophysiological mechanisms of selective auditory attention are based on neural gain when the notches are clearly wider than the critical band, and on neural tuning (e.g. gain+sharpenig) when the notches are within the critical band. In addition, the experimental data of the current study indicate that the frequency-specific attentional selection could have distinct neural mechanisms. The temporal analysis of the current data (Kauramäki et al., 2012) showed differences in the evolution of attention-related modulation between N100m and SF components. The modulation of the transient N100m components showed up immediately after attention was directed to the frequency discrimination task, whereas on sustained fields the modulation evolved latently and strengthened during the progression of auditory task. Moreover, there were also locational differences in the attention-related neural activity enhancements between the N100m and SF components. The spatial analysis of the current data (ECD and MNE estimates (Kauramäki et al., 2012)) showed that the enhanced neural activity was centered generally in the secondary auditory cortices, whose anatomical locations were in planum temporale at superior temporal gyrus. The locational differences were that N100m component sources were localized in the posterior areas of the secondary auditory cortices, whereas SF component sources were distributed rather to the medial regions of the left secondary auditory cortex (Kauramäki et al., 2012).

The temporal and locational differences between components can imply that there exist separate attention-related neural mechanisms (e.g. gain versus sharpening), or the frequency-specific selection is executed in separate neural levels. However, the findings of the current study can be explained also by increased inhibitory activity that is needed to attenuate the masker-induced suppression in the tonotopic auditory areas. The response latencies have been showed to be 200 ms or longer in experiments where the selective attention was used to exclude the irrelevant or distracting sensory inputs (Giard et al., 2000; Chait et al., 2010). In the current study the latencies of N100m components were about 200 ms with narrow notches ($\leq \pm 200$), which can signify that the attention-related mechanisms on the late N100m components ($\sim$ 200 ms) and on the SF components is the same. Thus, the attention-related enhancements after 200 ms post-onset could result from increased inhibitory activity that is needed to attenuate the masker-induced stimulation in the tonotopic auditory areas. The potential neural level mechanism can be the lateral inhibition (i.e. center-excitation and surround inhibition). This mechanism can reduce the overlap between neural populations that represent different frequencies, and thus, can explain the amplitude enhancements of both mechanisms – gain and sharpening.

[3] Correlation analysis of the current data was published by Kauramäki et al. (2012).

## 5.3 Conclusion

We humans are capable to listen to task-relevant auditory stream in conditions where the acoustic scene is distracted with simultaneous background noise. The selection of the attended tones succeeds even if the listening condition seems to be beyond the physical limits of hearing; in noise conditions, where the distractors are so close to the attended ones that they tend to stimulate the same neural ensembles at tonotopic auditory cortex. The aforementioned fact indicates that selective attention on frequency can cause short-term adjustments on tonotopic cortex that improve frequency selectivity.

The goal of the current study was of to explore the short-term neural mechanisms of selective auditory attention in conditions, where the notched-noise maskers were narrowed within the critical bandwidth of target tones ($\leq \pm 160$ Hz). In addition, the study aimed to show evidence of the neural tuning mechanism (gain + selectivity increase) in the narrow-notch conditions. The presumption was that the selectivity increase would be a more crucial mechanism in conditions where the frequency domain of the masker is overlapping with the critical bandwidth of target frequency.

The current data showed that the standard evoked auditory N100m and sustained field strengths were stronger under selective auditory attention. The effect was most robust at 200–500 ms post-onset when the notches were within the critical band ($\leq \pm 160$ Hz). Moreover, an interesting discovery was that the effect was stronger in the left hemisphere, especially with the thinnest notches ($\pm 150$ Hz – 0 Hz). The results confirmed (hypothesis c 2.6, p. 48) that the modulatory effect of selective attention was more powerful in the narrow notch conditions ($\leq \pm 160$ Hz). In addition, the results indicate that the neurophysiological mechanisms of selective auditory attention are based on bilateral neural gain when the notches are clearly wider than the critical band and on neural tuning (gain + selectivity increase) when the notches are within the critical band. Moreover, the results suggest that the left auditory cortex has a more active role in conditions where the segregation of relevant sounds from noise requires very sharp filtration.

The results imply that the frequency-specific attentional selection is executed in separate phases of even in separate neural levels; for example, a bilateral gain at 100–200 ms post-onset and a left-lateralized sharpening at 200–500 ms post-onset. However, the findings of the current study can also be explained by increased inhibitory activity that is needed to attenuate the masker-induced suppression in the tonotopic auditory areas. The potential neural level mechanism, lateral inhibition can give reasons, how the contrast between neural populations that represent different frequencies can be sharpened. Ultimately, this would reduce the overlap between neural representation (signal versus noise).

The findings are based on non-invasive measurements that represent population level neural effects. The tone-evoked responses arise from a combination of inhibitory and excitatory modulations that occur in the tonotopic auditory cortices. Thus, an interesting topic for further research would be a notched-noise single cell measurement in humans (notch width $\leq$ CB), which would give more detailed information of the microscopic neural level mechanisms of the selective auditory attention. This would provide also consistency to

the findings of the current study.

# Bibliography


Aben B, Stapert S, Blokland A (2012) About the distinction between working memory and short-term memory. *Frontiers in psychology* 3:301.

Ahveninen J, Hämäläinen M, Jääskeläinen I, Ahlfors S, Huang S, Lin F, Raij T, Sams M, Vasios C, Belliveau J (2011) Attention-driven auditory cortex short-term plasticity helps segregate relevant sounds from noise. *Proceedings of the National Academy of Sciences* 108:4182.

Ahveninen J, Jääskeläinen I, Raij T, Bonmassar G, Devore S, Hämäläinen M, Levänen S, Lin F, Sams M, Shinn-Cunningham B et al. (2006) Task-modulated "what" and "where" pathways in human auditory cortex. *Proceedings of the National Academy of Sciences* 103:14608.

Alberti P (2001) The anatomy and physiology of the ear and hearing. *Occupational Exposure to Noise: Evaluation, Prevention, and Control* pp. 53–62.

Alcaini M, Giard M, Echallier J, Pernier J (1995) Selective auditory attention effects in tonotopically organized cortical areas: A topographic erp study. *Human Brain Mapping* 2:159–169.

Altmann C, Henning M, Döring M, Kaiser J (2008) Effects of feature-selective attention on auditory pattern and location processing. *NeuroImage* 41:69–79.

Arnsten A, Berridge C, McCracken J (2009) The neurobiological basis of attention-deficit/hyperactivity disorder. *Primary Psychiatry* 16:47.

Arthur D, Lewis P, Medvick P, Flynn E (1991) A neuromagnetic study of selective auditory attention. *Electroencephalography and Clinical Neurophysiology* 78:348–360.

Ashmore J (2008) Hearing.

Ashmore J, Gale J (2000) The cochlea. *Current Biology* 10:325–327.

Baars B, Gage N (2010) *Cognition, brain, and consciousness: Introduction to cognitive neuroscience* Academic Press.

Barton R, Harvey P (2000) Mosaic evolution of brain structure in mammals. *Nature* 405:1055–1058.

Bidet-Caulet A, Fischer C, Besle J, Aguera P, Giard M, Bertrand O (2007) Effects of

selective attention on the electrophysiological representation of concurrent sounds in the human auditory cortex. *The Journal of Neuroscience* 27:9252.

Broadbent D (1958) *Perception and communication* Oxford University Press.

Brownell W, Oghalai J (2009) Cochlear biophysics. *Ballenger's Otorhinolaryngology: Head and Neck Surgery* p. 101.

Budd T, Barry R, Gordon E, Rennie C, Michie P (1998) Decrement of the n1 auditory event-related potential with stimulus repetition: Habituation vs. refractoriness. *International Journal of Psychophysiology* 31:51–68.

Campbell N, Reece J (2009) *Biology with mastering biology: International version* Pearson.

Chait M, de Cheveigné A, Poeppel D, Simon J (2010) Neural dynamics of attending and ignoring in human auditory cortex. *Neuropsychologia* .

Cherry E (1953) Some experiments on the recognition of speech, with one and with two ears. *Journal of the Acoustical Society of America* 25:975–979.

Corbetta M, Shulman G (2002) Control of goal-directed and stimulus-driven attention in the brain. *Nature reviews neuroscience* 3:201–215.

Cowan N (1998) *Attention and memory: An integrated framework* Oxford University Press.

da Silva F (2004) Functional localization of brain sources using eeg and/or meg data: Volume conductor and source models. *Magnetic Resonance Imaging* 22:1533–1538.

Darwin C, Turvey M, Crowder R (1972) An auditory analogue of the sperling partial report procedure: Evidence for brief auditory storage. *Cognitive Psychology* 3:255–267.

Dempster A, Laird N, Rubin D (1977) Maximum likelihood from incomplete data via the em algorithm. *Journal of the Royal Statistical Society. Series B (Methodological)* pp. 1–38.

Desimone R, Duncan J (1995) Neural mechanisms of selective visual attention. *Annual review of neuroscience* 18:193–222.

Driver J, Spence C (1998) Crossmodal attention. *Current Opinion in Neurobiology* 8:245–253.

Fletcher H (1940) Auditory patterns. *Reviews of Modern Physics* 12:47.

Freberg L (2009) *Discovering biological psychology* Wadsworth Pub Co.

Gazzaniga M, Ivry R, Mangun G (2002) *Cognitive neuroscience, Second Edition* W. W. Norton & Company.

Giard M, Fort A, Mouchetant-Rostaing Y, Pernier J (2000) Neurophysiological mechanisms of auditory selective attention in humans. *Frontiers in Bioscience* 5:d84.

Glasberg B, Moore B (1990) Derivation of auditory filter shapes from notched-noise data. *Hearing Research* 47:103–138.

Grady C, Van Meter J, Maisog J, Pietrini P, Krasuski J, Rauschecker J (1997) Attention-related modulation of activity in primary and secondary auditory cortex. *Neuroreport* 8:2511–2516.

Green D, Swets J (1988) *Signal detection theory and psychophysics* Peninsula Publishing.

Hansen J, Hillyard S (1980) Endogeneous brain potentials associated with selective auditory attention. *Electroencephalography and Clinical Neurophysiology* 49:277–290.

Hari R, Aittoniemi K, Järvinen M, Katila T, Varpula T (1980) Auditory evoked transient and sustained magnetic fields of the human brain localization of neural generators. *Experimental Brain Research* 40:237–240.

Hari R, Mäkelä J (1988) Modification of neuromagnetic responses of the human auditory cortex by masking sounds. *Experimental Brain Research* 71:87–92.

Hickok G, Houde J, Rong F (2011) Sensorimotor integration in speech processing: computational basis and neural organization. *Neuron* 69:407–422.

Hillyard S, Hink R, Schwent V, Picton T (1973) Electrical signs of selective attention in the human brain. *Science* 182:177.

Hillyard S, Vogel E, Luck S (1998) Sensory gain control (amplification) as a mechanism of selective attention: electrophysiological and neuroimaging evidence. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences* 353:1257–1270.

Hocherman S, Benson D, Goldstein M, Heffner H, Hienz R (1976) Evoked unit activity in auditory cortex of monkeys performing a selective attention task. *Brain Research* 117:51–68.

Hodgkin A, Huxley A (1952) A quantitative description of membrane current and its application to conduction and excitation in nerve. *The Journal of Physiology* 117:500.

Humphries C, Liebenthal E, Binder J (2010) Tonotopic organization of human auditory cortex. *NeuroImage* 50:1202–1211.

Hämäläinen M, Hari R (2002) Magnetoencephalographic characterization of dynamic brain activation: Basic principles and methods of data collection and source analysis. *Brain Mapping: The Methods. London, Elsevier* pp. 227–253.

Hämäläinen M, Hari R, Ilmoniemi R, Knuutila J, Lounasmaa O (1993) Magnetoencephalography – theory, instrumentation, and applications to noninvasive studies of the working human brain. *Reviews of Modern Physics* 65:413–497.

Inui K, Okamoto H, Miki K, Gunji A, Kakigi R (2006) Serial and parallel processing in the human auditory cortex: A magnetoencephalographic study. *Cerebral Cortex* 16:18.

Javitt DC, Sweet RA (2015) Auditory dysfunction in schizophrenia: integrating clinical and basic features. *Nature Reviews Neuroscience* 16:535–550.

Jääskeläinen I, Ahveninen J, Belliveau J, Raij T, Sams M (2007) Short-term plasticity in auditory cognition. *Trends in Neurosciences* 30:653–661.

Jääskeläinen I, Ahveninen J, Bonmassar G, Dale A, Ilmoniemi R, Levänen S, Lin F, May P, Melcher J, Stufflebeam S et al. (2004) Human posterior auditory cortex gates novel sounds to consciousness. *Proceedings of the National Academy of Sciences of the United States of America* 101:6809.

Kandel E, Schwartz J, Jessell T, Mack S, Dodd J (2000) *Principles of neural science*, Vol. 4 McGraw-Hill Medical.

Kanno A, Nakasato N, Fujiwara S, Yoshimoto T (1996) Right hemispheric dominancy in the auditory evoked magnetic fields for pure-tone stimuli. *No To Shinkei – Brain and Nerve* 48:240.

Kant I, Kehrbach K (1910) *Kritik der reinen Vernunft* A. Kröner.

Karjalainen M (2009) *Kommunikaatioakustiikka* Multiprint Oy.

Kauramäki J, Jääskelainen I, Sams M (2007) Selective attention increases both gain and feature selectivity of the human auditory cortex. *PLoS One* 2.

Kauramäki J, Jääskeläinen I, Hänninen J, Auranen T, Nummenmaa A, Lampinen J, Sams M (2012) Two-stage processing of sounds explains behavioral performance variations due to changes in stimulus contrast and selective attention: An meg study. *PloS One* 7:e46872.

Knudsen E (2007) Fundamental components of attention. *Neuroscience* 30:57.

Konishi T, Hamrick P, Walsh P (1978) Ion transport in guinea pig cochlea: I. potassium and sodium transport. *Acta Oto-Laryngologica* 86:22–34.

Kotchoubey B (2006) Event-related potentials, cognition, and behavior: A biological approach. *Neuroscience & Biobehavioral Reviews* 30:42–65.

Lachter J, Forster K, Ruthruff E (2004) Forty-five years after broadbent (1958): Still no identification without attention. *Psychological Review* 111:880.

Levitt H et al. (1971) Transformed up-down methods in psychoacoustics. *Journal of the Acoustical Society of America* 49:467–477.

Liegeois-Chauvel C, de Graaf J, Laguitton V, Chauvel P (1999) Specialization of left auditory cortex for speech perception in man depends on temporal coding. *Cerebral cortex* 9:484–496.

Luck S (2005) *An introduction to the event-related potential technique* Cambridge, Mass.: The MIT Press.

Malmivuo J, Plonsey R (1995) *Bioelectromagnetism: Principles and applications of bioelectric and biomagnetic fields* Oxford University Press.

Martin J (2003) *Neuroanatomy: Text and atlas* McGraw-Hill Medical.

Matthews P, Jezzard P (2004) Functional magnetic resonance imaging. *Journal of Neurology, Neurosurgery & Psychiatry* 75:6–12.

May P, Tiitinen H (2010) Mismatch negativity (mmn), the deviance-elicited auditory deflection, explained. *Psychophysiology* 47:66–122.

Michel C, Koenig T, Brandeis D, Gianotti L, Wackermann J, Corporation E (2009) *Electrical neuroimaging* Cambridge University Press.

Moore B (1995) *Hearing (Handbook of Perception and Cognition, Second Edition)* Academic Press.

Moore B (2008) Basic auditory processes involved in the analysis of speech sounds. *Philosophical Transactions of the Royal Society B: Biological Sciences* 363:947–963.

Murray S (2005) Attention and changes in neural selectivity. *Neurobiology of Attention* pp. 485–489.

Niiniluoto I, Koskinen H (1995) Kognitiivisesta naturalismista naturalistiseen kognitioon. *Helsinki: Helsingin yliopiston filosofian ja psykologian laitosten järjestämä kotimainen kollokvio 8.11.1995* 8.

Näätänen R (1990) The role of attention in auditory information processing as revealed by event-related potentials and other brain measures of cognitive function. *Behavioral and Brain Sciences* .

Näätänen R, Gaillard A, Mäntysalo S (1978) Early selective-attention effect on evoked potential reinterpreted. *Acta Psychologica* 42:313–329.

Näätänen R, Picton T (1987) The n1 wave of the human electric and magnetic response to sound: A review and an analysis of the component structure. *Psychophysiology* 24:375–425.

Okamoto H, Stracke H, Bermudez P, Pantev C (2011) Sound processing hierarchy within human auditory cortex. *Journal of Cognitive Neuroscience* 23:1855–1863.

Okamoto H, Stracke H, Ross B, Kakigi R, Pantev C (2007a) Left hemispheric dominance during auditory processing in a noisy environment. *BMC Biology* 5:52.

Okamoto H, Stracke H, Wolters C, Schmael F, Pantev C (2007b) Attention improves population-level frequency tuning in human auditory cortex. *The Journal of Neuroscience* 27:10383–10390.

Pantev C, Bertrand O, Eulitz C, Verkindt C, Hampson S, Schuierer G, Elbert T (1995) Specific tonotopic organizations of different areas of the human auditory cortex revealed by simultaneous magnetic and electric recordings. *Electroencephalography and Clinical Neurophysiology* 94:26–40.

Pantev C, Eulitz C, Elbert T, Hoke M (1994) The auditory evoked sustained field: ori-

gin and frequency dependence. *Electroencephalography and Clinical Neurophysiology* 90:82–90.

Parkkonen L (2010) Instrumentation and data preprocessing. *MEG: An Introduction to Methods* pp. 24–64.

Patterson R (1986) Auditory filters and excitation patterns as representations of frequency resolution. *Frequency Selectivity in Hearing* pp. 123–177.

Petkov C, Kang X, Alho K, Bertrand O, Yund E, Woods D (2004) Attentional modulation of human auditory cortex. *Nature Neuroscience* 7:658–663.

Picton T, Woods D, Proulx G (1978) Human auditory sustained potentials. i. the nature of the response. *Electroencephalography and Clinical Neurophysiology* 45:186–197.

Posner M, Petersen S (1990) The attention system of the human brain. *Annual Review of Neuroscience* 13:25–42.

Posner M, Rothbart M (2007) Research on attention networks as a model for the integration of psychological science. *Annual Review of Psychology* 58:1–23.

Purves D, Augustine G, Fitzpatrick D, Katz L, LaMantia A, McNamara J, Williams S (2004) *Neuroscience, 3rd edition* Sinauer Associates, Inc.

Ramírez R, Wipf D, Baillet S (2010) Neuroelectromagnetic source imaging of brain dynamics. *Computational Neuroscience* pp. 127–155.

Rauschecker J, Romanski L (2011) Auditory cortical organization: Evidence for functional streams. *The Auditory Cortex* pp. 99–116.

Rif J, Hari R, Hamalainen M, Sams M (1991) Auditory attention affects two different areas in the human supratemporal cortex. *Electroencephalography and Clinical Neurophysiology* 79:464–472.

Roelfsema PR, Lamme VA, Spekreijse H (1998) Object-based attention in the primary visual cortex of the macaque monkey. *Nature* 395:376–381.

Rosburg T, Trautner P, Korzyukov O, Boutros N, Schaller C, Elger C, Kurthen M (2004) Short-term habituation of the intracranially recorded auditory evoked potentials p50 and n100. *Neuroscience Letters* 372:245–249.

Rossing T, Moore F, Wheeler P (2002) *The science of sound* Addison-Wesley Reading. MA.

Salmi J, Rinne T, Koistinen S, Salonen O, Alho K (2009) Brain networks of bottom-up triggered and top-down controlled shifting of auditory attention. *Brain research* 1286:155–164.

Sams M, Salmelin R (1994) Evidence of sharp frequency tuning in the human auditory cortex. *Hearing Research* 75:67–74.

Schreiner C, Froemke R, Atencio C (2011) Spectral processing in auditory cortex. *The Auditory Cortex* pp. 275–308.

Seikel J, King D, Drumright D (2009) *Anatomy and physiology for speech, language, and hearing* Delmar Pub.

Singh K (2006) Magnetoencephalography. *Methods in Mind* pp. 291–326.

Talavage T, Sereno M, Melcher J, Ledden P, Rosen B, Dale A (2004) Tonotopic organization in human auditory cortex revealed by progressions of frequency sensitivity. *Journal of Neurophysiology* 91:1282.

Woldorff M, Gallen C, Hampson S, Hillyard S, Pantev C, Sobel D, Bloom F (1993) Modulation of early sensory processing in human auditory cortex during auditory selective attention. *Proceedings of the National Academy of Sciences* 90:8722.

Woldorff M, Hillyard S (1991) Modulation of early auditory processing during selective listening to rapidly presented tones. *Electroencephalography and Clinical Neurophysiology* 79:170–191.

Woods D, Alho K, Algazi A (1994) Stages of auditory feature conjunction: An event-related brain potential study. *Journal of Experimental Psychology: Human Perception and Performance* 20:81.

Yantis S (2008) The neural basis of selective attention: Cortical sources and targets of attentional modulation. *Current Directions in Psychological Science* 17:86.

Zani A, Proverbio A, Posner M (2003) *The cognitive electrophysiology of mind and brain* Academic Press, Elsevier.

Zatorre R, Belin P, Penhune V (2002) Structure and function of auditory cortex: Music and speech. *Trends in Cognitive Sciences* 6:37–46.

Zwicker E, Terhardt E (1980) Analytical expressions for critical-band rate and critical bandwidth as a function of frequency. *The Journal of the Acoustical Society of America* 68:1523.